Prepared for submission to JHEP JLAB-THY-22-..., YITP-SB-2022-...

Is infrared-collinear safe information all you need for jet classification?

Dimitrios Athanasakos,[1,2] **Andrew J. Larkoski,**[3] **James Mulligan,**[4,5] **Mateusz Płoskoń,**[4] **Felix Ringer**[1,2,6,7]

[1] *C.N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, NY 11794, USA*

[2] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*

[3] *Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*

[4] *Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[5] *Physics Department, University of California, Berkeley, CA 94720, USA*

[6] *Department of Physics, Old Dominion University, Norfolk, VA 23529, USA*

[7] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*

E-mail: dimitrios.athanasakos@stonybrook.edu, larkoski@ucla.edu, james.mulligan@berkeley.edu, mploskon@lbl.gov, fmringer@jlab.org

Abstract: Machine learning-based jet classifiers are able to achieve impressive tagging performance in a variety of applications in high energy and nuclear physics. However, it remains unclear in many cases which aspects of jets give rise to this discriminating power, and whether jet observables that are calculable in perturbative QCD such as those obeying infrared-collinear (IRC) safety serve as sufficient inputs. In this article, we introduce a new classifier, Jet Flow Networks (JFNs), in an effort to address the question of whether IRC unsafe information provides additional discriminating power in jet classification. JFNs are permutation-invariant neural networks (deep sets) that take as input the kinematic information of reconstructed subjets. The subjet radius serves as a tunable hyperparameter, enabling the sensitivity to soft emissions and nonperturbative effects to be gradually increased as the subjet radius is decreased. We demonstrate the performance of JFNs for quark vs. gluon and QCD vs. Z jet tagging. For small subjet radius, the performance of JFNs is comparable to the IRC-unsafe Particle Flow Networks (PFNs), demonstrating that infrared-collinear unsafe information is not necessary to achieve strong discrimination. As the subjet radius is increased, the performance of the JFNs remains essentially unchanged until physical thresholds that we identify are crossed. For relatively large subjet radii, we show that the JFNs may offer an increased model independence with a modest tradeoff in performance compared to classifiers that use the full particle information of the jet. Our results shed new light onto how machines learn patterns in high-energy physics data.[1]

[1] Title inspired by Ref. [1].

Contents

1 Introduction

Jets are highly energetic and collimated groups of particles observed in the detectors of high-energy scattering experiments such as the Large Hadron Collider (LHC) [2–4]. Jets arise from the fragmentation of highly energetic quarks and gluons, which themselves can arise from the decay of unstable particles such as the Higgs boson. Classifying the origins of jets, such as quark vs. gluon initiated jets [5, 6] or QCD vs. boosted Z/W jets [7], **(want more refs here, or perhaps just review articles? –ajl)** is crucial to disentangle the various processes occurring at collider experiments and perform searches for physics beyond the Standard Model.

Jet classification algorithms have been developed both based on multivariate combinations of jet substructure observables, as well as using machine learning methods. Machine learning based jet classifiers significantly outperform traditional multivariate jet taggers based on a limited number of observables, since they are able to leverage the full information in the jet [8]. However, machine learning based classifiers often have the drawback that they are not calculable by analytical methods. Efforts to address this have been an active area of research, such as enforcing Infrared-Collinear (IRC) safety in the network architecture [9] or by varying the amount of information provided as input to neural networks in order to deduce the relevant aspects responsible for the classification power [10], e.g. by employing Lasso regression [11].

In order to increase the interpretability of machine learning based classifiers, a complete IRC-safe basis of jet substructure observables was introduced in Refs. [8, 12, 13] based on N-subjettiness observables [14, 15]. These observables capture the momentum and relative

angles of emissions inside the jet. The set of N-subjettiness observables is then used as input to a machine learning algorithm for jet classification. While the complete basis of IRC-safe observables is large ($3M-4$ for M particles in the jet), it was found that the performance of classifiers saturates quickly with a relatively small number of observables. Another set of observables, Energy Flow Polynomials (EFPs), was developed as a linear and IRC-safe basis of jet substructure observables in Ref. [9].

Interestingly, it was found that while the performance of classifiers based on complete sets of observables saturates, there remains a performance gap between classifiers with IRC-safe inputs (Sudakov safe classifiers) and IRC-unsafe classifiers that make use of the full information content of the particles inside the jet. Examples of such IRC-unsafe classifiers include architectures based on deep sets [16], point clouds [17] and transformers [18]. This performance gap has been observed for a variety of jet classification tasks, including quark vs. gluon tagging [16, 19], QCD vs. W and H jets [20], and pp vs. AA jets [10]. Several efforts have been made to quantify the gap, with the aim to gain new insights into fundamental QCD dynamics [**?**]. There are several possible explanations for the observed performance gap:

- IRC-unsafe classifiers may be able to make use of the very soft information content of jets, which is difficult to access with IRC-safe observables.
- IRC-unsafe classifiers such as PFNs take as input the exact position information of the particles inside the jet, whereas IRC-safe observables can only capture the information of relative distances. It is possible that existing machine learning algorithms can make more efficient use of position information.
- The specific form of the IRC-safe observables may not be optimal for classification tasks and there may be other sets of observables that could perform better.

With this question in mind, we introduce in this work a new machine learning-based jet classifier, Jet Flow Networks (JFNs)[1], which take as input the energy and position of reclustered subjets instead of individual particles. JFNs allow for soft and collinear emissions to be clustered into subjets making the input IRC-safe and the resulting classifier generally Sudakov safe [21, 22]. However, different than the N-subjettiness or the EFP basis of observables, position information is used instead of having (indirectly) access only to relative distances between emissions (or subjets) inside the jet. We note that we do not consider quark flavor tagging in this work which requires nonperturbative information, see e.g. Refs. [23–25].

JFNs are closely related to Particle Flow Networks (PFNs) [16] and Energy Flow Networks (EFNs) [16], which will be elaborated on in section 3. In the limit of a vanishing subjet radius where every subjet contains only a single hadron, JFNs are identical to PFNs. The radius of the reclustered subjets in JFNs can be used to dial in nonperturbative information allowing for a smooth transition to IRC-unsafe classifiers. As such, JFNs complement the existing family of permutation-invariant networks in particle physics. As

[1] In analogy to Particle Flow Networks (PFNs) [16].

for PFNs and EFNs, we will utilize machine learning algorithms for JFNs based on a permutation invariant deep set architecture [16, 26–28].

The main result of our work will be to show that the JFNs based on IRC-safe subjets achieve the same classification performance as PFNs for a finite, non-zero subjet radius. The exact value of the subjet radius where the PFN performance is matched depends on the classification task at hand. Therefore, different than the classifiers based on complete IRC-safe sets of observables, JFNs constitute a "gapless" classifier indicating that the very soft aspects of jets are in fact not relevant for typical classification tasks at collider experiments. This answers in part the question about the features that are relevant for the performance of classifiers in high-energy physics. Throughout this work, PFNs are taken as a reference, but other permutation invariant classifiers such as GNNs and transformers could equally well be trained on particles or reclustered subjets.

In addition, shedding light on the role of IRC-safe information, JFNs allow for new insights into the physics of jet tagging and may lead to various future applications at high-energy collider experiments. By studying the performance of the JFNs as a function of the subjet radius and the jet transverse momentum, we are able to identify the relevant physical scales of different classification tasks. For example, for QCD vs. Z-jet tagging, we find that ... (Insert final conclusion here). Second, we explore the generalization capability of JFNs to unseen data, which is crucial when deploying a classifier trained on simulations to experimental data. Due to the clustering of collinear and soft emissions into subjets, the resulting JFNs are relatively insensitive to the detailed modeling of the infrared (IR) physics that is often poorly understood. This raises the possibility to use JFNs to trade performance for generalizability by adjusting the number of reconstructed subjets. Lastly, we expect that subjets can be measured well in heavy-ion collisions despite the large fluctuating background. See Ref. [29] for recent measurements of the energy spectrum of inclusive and leading subjets by the ALICE Collaboration. See also Ref. [30].

The remainder of this paper is organized as follows. In section 2, we introduce the subjet basis and discuss differences between inclusive and exclusive subjet reconstruction algorithms. In section 3, we introduce the permutation invariant machine learning algorithms that take the kinematic information of the reconstructed subjets as input and in section 4, we briefly discuss the data sets used for different classification tasks used in this work. In section 5, we present numerical results for the classification performance of JFNs for quark vs. gluon and QCD vs. Z jets. In particular, we show that JFNs match the PFN performance for a finite subjet radius. Based on these results, we describe in section 6 that the machine learning algorithm is sensitive to different physical scales, which it can effectively learn. In section 7, we investigate the tradeoff between performance and generalizability of the JFNs. In section 8, we draw conclusions and present an outlook.

2 The subjet basis

In this section, we describe the reconstruction of subjets that will serve as the input to the machine learning classifier. The initial jet is identified using the anti-k_T algorithm [31] and jet radius parameter R. In order to utilize the substructure of jets, we then recluster the

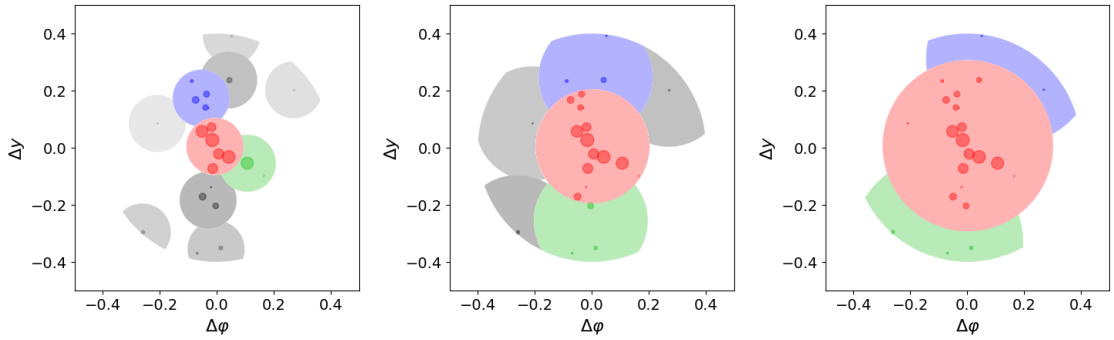

Figure 1: Illustration of a QCD jet with $p_T = 100$ GeV and radius parameter $R = 0.4$ reclustered into subjets for subjet radii $r = 0.1$ (left), $r = 0.2$ (middle), and $r = 0.3$ (right). We use the inclusive anti-k$_T$ algorithm to identify the initial jet and the subjets. Particles are represented by small filled circles with radii proportional to the particle transverse momentum in the Δy vs. $\Delta\varphi$ plane, where $\Delta\varphi = \varphi^{\text{particle|subjet}} - \varphi^{\text{jet}}$ is the azimuthal angle with respect to the jet axis and $\Delta y = \Delta y^{\text{particle|subjet}} - \Delta y^{\text{jet}}$ is rapidity distance to the jet axis. Subjets are shown with larger colored areas where red marks the leading subjet, green marks the second leading subjet, blue marks the third leading jet, and shades of gray represent subjets with lower momentum fraction z with intensity proportional to z.

jet constituents into subjets. We consider two approaches for the subjet reconstruction: inclusive anti-k$_T$ subjets and the exclusive k$_T$ subjets [32, 33]. In both cases soft and collinear emissions are first clustered into subjets, making the input to the IRC safe. In this sense, subjets serve as a useful tool for throttling or controlling the input data to the machine in a way that is theoretically interpretable in perturbative QCD.

First, we consider inclusive subjets reconstructed with the anti-k$_T$ algorithm and a fixed jet radius $r < R$. This approach fixes the maximally allowed size of the reconstructed subjets but the number of subjets varies for each jet. We illustrate the distribution of subjets in the η-ϕ plane for three different subjet radii in Fig. 1. As r is increased, the central subjet contains a large fraction of particles. Second, we consider subjets reconstructed with the exclusive k_T algorithm. Particles are clustered with the k_T algorithm until a fixed number of subjets N is obtained. Different than in the case of inclusive subjets, the number of identified subjets is fixed but their size varies jet-by-jet. The N subjets span the full information content of the N most resolved emissions inside the jet analogous to the N-subjettiness basis developed in Refs. [8, 12, 13]. An alternative approach to the exclusive k_T algorithm is to identify subjets with the XCone algorithm [34]. We leave the exploration of this algorithm for future work. By taking the small-r (inclusive subjets) or large-N limit (exclusive subjets), we can study the transition to the nonperturbative regime where eventually, every subjet only contains a single hadron.

To illustrate the qualitative differences between the two reconstruction methods discussed above, we show as an example the longitudinal momentum distributions of subjets $z = p_T^{\text{subjet}}/p_T$ in Fig. 2 separately for quark and gluon jets. Here p_T denotes the initial jet transverse momentum and p_T^{subjet} the longitudinal subjet momentum using either the inclu-

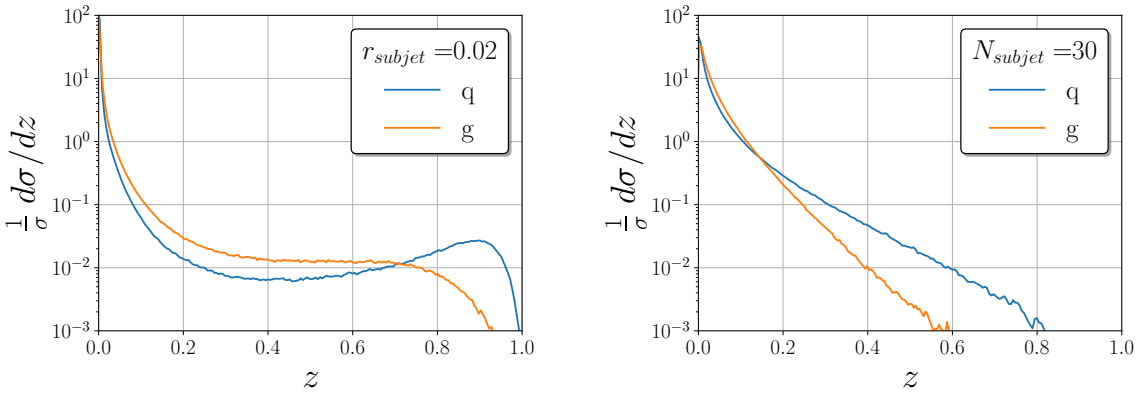

Figure 2: The longitudinal momentum distribution of inclusive subjets $z = p_T^{\text{subjet}}/p_T$ originating from either a quark (blue) or a gluon (orange) jet. We show the distributions for inclusive subjet clustering with $r = 0.02$ (left) and for exclusive clustering with a fixed number of $N = 30$ subjets (right), which yields a comparable average number of subjets.

sive or exclusive reconstruction method. As an example, we choose $N = 30$ for the exclusive reconstruction of subjets and $r = 0.02$ for inclusive subjets, which yields a comparable average number of subjets. We observe that the two methods lead to qualitatively different spectra. The inclusive subjet spectrum exhibits a peak (quarks) or plateau (gluons) for intermediate to large values of z. In contrast, the spectrum for exclusive subjets only peaks at small values of z and falls off steeply for $z \to 1$. This is due to the fact that for exclusive clustering the k_T algorithm is used, where "soft" hadrons are clustered first. Only at the end "hard" emissions are combined, making it unlikely to find a subjet with $z \to 1$ for a fixed value of N. We note that for inclusive subjets, the z-distributions are qualitatively the same for both the anti-k_T and k_T algorithms. The longitudinal momentum spectrum for inclusive subjets was calculated within perturbative QCD up to next-to-leading logarithmic (NLL) accuracy. See Refs. [35–38]. This close connection to first-principles calculations may allow for an increased understanding of machine learning algorithms in QCD.

From the identified subjets, the kinematic information (z_i, η_i, ϕ_i) of the each subjet is used as input to the classifiers discussed below. In the limit that $r \to 0$ (inclusive subjets) or $N \to \infty$ (exclusive subjets), the subjet basis becomes equivalent to the set of particle four-vectors of the jet, and the classifier can make use of the full information content of the jet. The subjet basis therefore provides a means to limit the information supplied to the classifier, by using $r > 0$ or $N < \infty$.

3 Jet Flow Networks (JFNs): Deep sets of subjets

In this section, we describe the permutation invariant neural networks that use the kinematic information of subjets as input to perform binary classification tasks. As introduced above, we refer to the machine learning architecture and the pre-processing step of clustering particles into subjets as JFNs.

	PFN [16]	JFN	EFN [16]
Input	particle 4-momenta	subjet 4-momenta	particle 4-momenta
Classifier	IRC unsafe	Sudakov safe	IRC safe

Table 1: Overview of different classifiers based on permutation invariant neural networks.

The reconstructed subjets discussed in the previous section do not have an inherent ordering. Therefore, permutation-invariant neural networks are a natural choice to perform classification tasks that take as input the kinematic information of subjets. In Refs. [26–28] deep sets were introduced as a permutation invariant neural network. In the context of particle physics deep sets were first discussed in Ref. [16] as Particle Flow Networks (PFNs) that take as input the information of individual particles. A permutation invariant classifier f, which takes as input the subjet four-momenta p_i satisfies $f(p_1, \ldots, p_N) = f(p_{\pi(1)}, \ldots, p_{\pi(N)})$. Here π denotes the permutation operator. Following Ref. [26], we can write the classifier f as

$$f(p_1, \ldots, p_n) = F\left(\sum_{i=1}^{N} \Phi_i(p_i)\right), \tag{3.1}$$

where F, Φ are neural networks and, as an intermediate step, we sum over all reconstructed subjets N. The first neural network $\Phi : \mathbb{R}^4 \to \mathbb{R}^l$ takes as input the individual subjet four momenta and maps it to an l-dimensional latent space. For massless subjets, we can write the individual four vectors in terms of (z_i, η_i, ϕ_i). Here z_i is the subjet's longitudinal momentum fraction, see Fig. 2, and (η_i, ϕ_i) denote its coordinates in the rapidity-azimuth plane. We note that further information can be included in the per-subjet mapping such as the jet mass or the jet charge [39, 40], analogous to e.g. particle identification (PID) for PFNs. We leave quantitative studies of the impact of these additional features for future work. The summation in Eq. (3.1) ensures that the classifier f is invariant under permutations of the input variables. The second neural network $F : \mathbb{R}^l \to \mathbb{R}$ is a map from the latent space where the summation operation is performed to the final classification score. Note that the classifier architecture in Eq. (3.1) can accommodate both a fixed number N of subjets (exclusive subjets) and input with variable length (inclusive subjets).

We refer to the deep set classifier based on subjets in Eq. (3.1) as JFNs. The JFNs are a family of classifiers due to the dependence on the continuous parameter r in the case of inclusive clustering or on N in the case of exclusive clustering, in which case the clustering is performed until N subjets remain. Since the JFN takes subjet information as input, the resulting classifier is generally Sudakov safe [22]. We summarize the different aspects of permutation invariant network architectures based on deep sets in table 1. Since the JFNs are Sudakov safe, they constitute an intermediate point between IRC-unsafe PFNs and IRC-safe EFNs. In the limit of $r \to 0$ (inclusive subjets) or the large-N limit (exclusive subjets), we recover the PFN classifier.

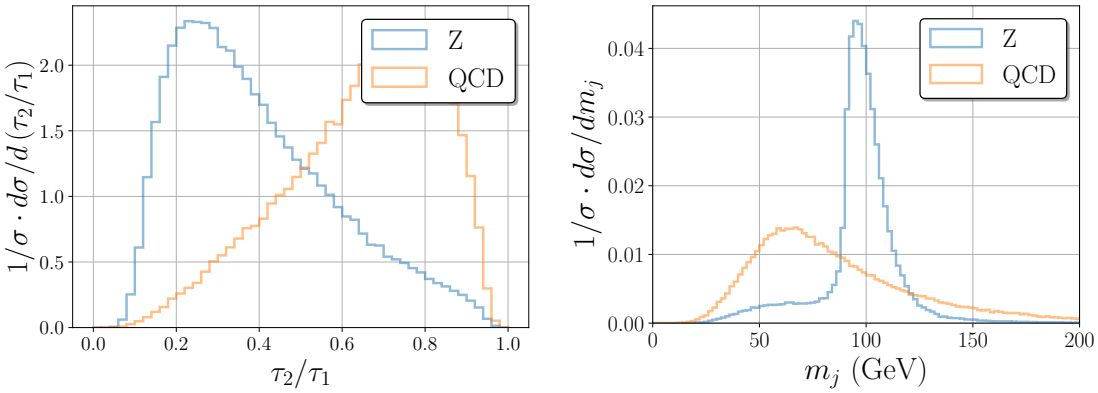

Figure 3: N-subjettiness ratio $\tau_2^{(1)}/\tau_1^{(1)}$ (left) and the jet mass distribution (right) for QCD and Z jets with $p_T = [500, 550]$ GeV. The N-subjettiness axes were identified using the one pass k_T clustering algorithm.

4 Data sets

In this work, we will consider JFNs for two exemplary binary classification tasks in high-energy physics. First, we consider quark vs. gluon jet classification and, second, QCD vs. Z jet classification. For the quark vs. gluon case, we make use of the data set in Ref. [41], which consists of 2M jets with transverse momentum $p_T = [500, 550]$ GeV, rapidity $|\eta| < 1.7$, jet radius parameter $R = 0.4$, and center-of-mass energy $\sqrt{s} = 14$ TeV. We will make use of both the data set generated with Pythia [42] and Herwig [43]. In order to explore the dependence on the jet transverse momentum, we also generate two additional data sets consisting of 250k jets each with transverse momentum $p_T = [300, 350]$ GeV and $[1000, 1050]$ GeV, respectively. The underlying processes are: $q\bar{q} \to Z(\to \nu\bar{\nu}) + g$ and $q\bar{q} \to Z(\to \nu\bar{\nu}) + (uds)$ analogous to Ref. [41]. For the QCD vs. Z-jet case, we generate 500k jets for three different bins of jet transverse momentum, [300, 350] GeV, [500, 550] GeV and [1000, 1100] GeV with a jet mass $m_j = [45, 135]$ GeV. The radius parameter is $R = 0.8$, the rapidity cut is $|\eta| < 1.7$ and the samples are generated using Pythia with the Monash2013 tune and Herwig at $\sqrt{s} = 14$ TeV. Jets arising from Z bosons are tagged according to the leading Z boson presence in the catchment area of the jets as extracted from the kinematics of the events at the particle level before hadronization with a Z-jet distance from the jet axis less than $R/2$. A similar tagging procedure is performed to differentiate between quark and gluon jets in the QCD sample. The tag is based on the leading parton within the catchment area of the jet. However, to strengthen the parton-jet association, we use parton-level kinematics injected into the hadron-level event using so-called ghost particles ($p_{\mathrm{T}} = 10^{-5}$ GeV/c) that do not affect jet reconstruction but allow for efficient tagging after the jet finding step.

The substructure of QCD jets is generally single-pronged, whereas the decay products of a Z cause the corresponding jets to have two prongs. The ratio of N-subjettiness observables [14, 15, 44, 45] is sensitive to the number of prongs inside a jet. In order to

define the N-subjettiness, a given number of N axes are identified inside the jet using the exclusive k_T algorithm. The N-subjettiness variables $\tau_N^{(\beta)}$ measure the radiation along these axes and are defined as

$$\tau_N^{(\beta)} = \frac{1}{p_T} \sum_{i \in \text{jet}} p_{Ti} \min \left\{ R_{1i}^\beta, R_{2i}^\beta, \ldots, R_{Ni}^\beta \right\} . \tag{4.1}$$

Here the p_{Ti} of each particle i is weighted by its distance R_{ji} to the closest axis j raised to the power $\beta > 0$, which is a tunable parameter. For single-prong jets, the variable τ_2 will peak at smaller values compared to τ_1, whereas for two-prong jets the variable τ_2 takes similar values compared to τ_1 (by construction, $\tau_{n+1} \leq \tau_n$). In the left panel of Fig. 3, we show the result for the ratio τ_2/τ_1, which shows the expected separation of the two jet samples, and the jet mass m_j distribution for QCD and Z jets (right panel). In order to capture the on-shell decays of the Z jets, we will restrict ourselves in the mass range $m_j = [75, 105]$ GeV for the analysis of the QCD vs. Z-jet discrimination problem. For all classification tasks, the training/validation/test split is 80%/10%/10%.

5 JFN performance: gapless jet classification

In this section, we will explore the performance of the JFNs and compare the results to PFNs, which are recovered as a limit of the JFNs where for $r \to 0$ every subjet contains only a single hadron. We consider two typical binary classification tasks in high-energy physics: quark vs. gluon jet and QCD vs. Z boson jet identification.

In order to implement the permutation-invariant neural networks, we parametrize the functions Φ and F in Eq. (3.1) in terms of DNNs, using the `EnergyFlow` package [16] with `Keras` [46]/`TensorFlow` [47]. For Φ we use two hidden layers with 100 nodes each and a latent space dimension of $d = 256$. For F we include three layers with 100 nodes each. For each dense layer, we use the ReLU activation function [48] and we use the softmax activation function for the final output layer of the classifier. We train the neural networks using the Adam optimizer [49] with learning rates ranging from 10^{-3} to 10^{-4}. We use the binary cross entropy loss function [50] and train for 50 epochs with a batch size of 256 and a patience parameter of 7. We find no significant changes in performance when changing the size or number of the layers, latent space dimension, learning rate, and batch size by factors of 2-5. Following Ref. [16], we perform a preprocessing step to simplify the training process: we use the rescaled momentum fractions z_i and center the rapidity and azimuthal angles η_i, ϕ_i of the particles in the jet with respect to the jet direction.

We quantify the performance of the different classifiers in terms of the Receiver Operator Characteristic (ROC) curve and the Area Under the Curve (AUC) of ROC curve. The ROC curve is the cumulative distribution function of the true positive rate vs. the false positive rate of a classifier as the decision threshold is varied. The AUC takes values between 0.5 and 1, where 0.5 (1) corresponds to a random (perfect) binary classifier. We estimate the statistical uncertainty of the AUC by training the deep sets five times for each choice of the subjet radius r and using the standard deviation as the uncertainty.

Fig. 4 (left) shows the JFN results for q vs. g jet discimination using inclusive subjet clustering. The top panel shows the AUC performance as we change the inclusive subjet

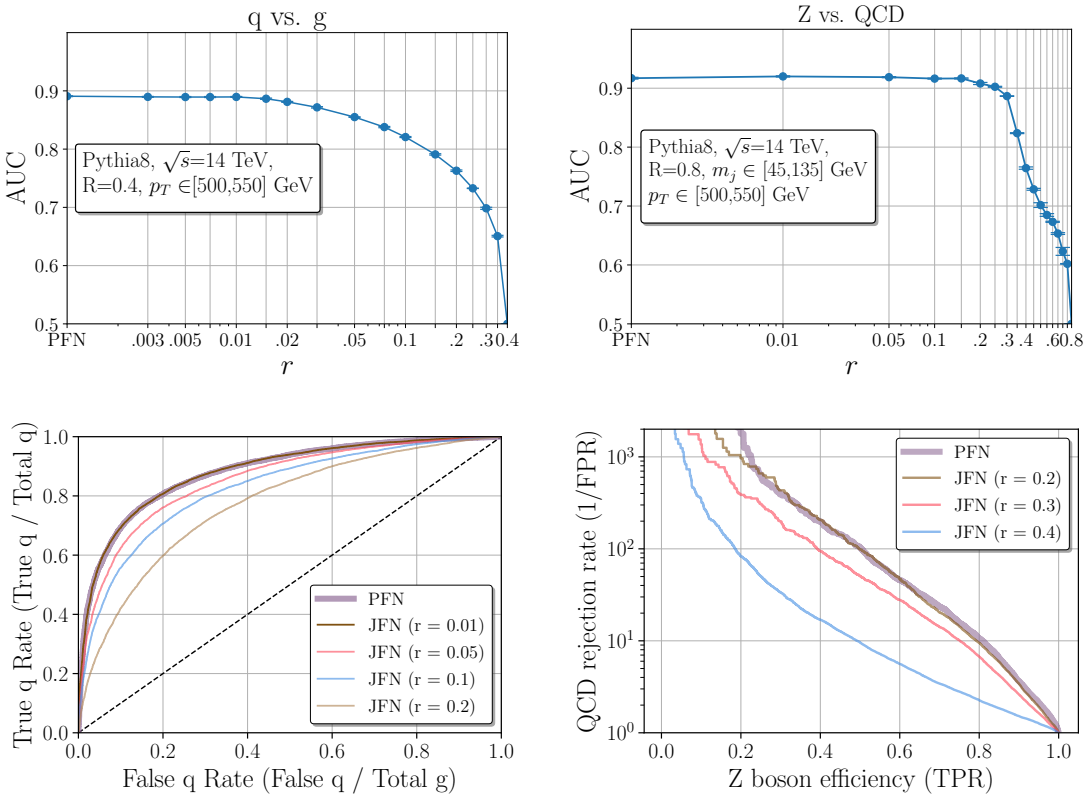

Figure 4: Top panel: AUC for quark vs. gluon jet (left) and QCD vs. Z jet tagging (right) using JFNs with different values of the (inclusive) subjet radius r. The PFN classifier is shown for reference at the leftmost value of r. Bottom panel: ROC curves for quark vs. gluon (left) and QCD vs. Z jets tagging using JFNs with different values of the (inclusive) subjet radius r for the same datasets as the upper panel.

radius r, and the bottom panel shows the ROC curves for several r. For comparison, we also show the result for PFNs. In the case of the AUC plot, we display the PFN classifier as the leftmost point on the r axis. As expected, we find that for the smallest subjet radii r the performance of the PFN is recovered. Strikingly, however, the performance of the JFN does not significantly diminish as r is increased for values of the subjet radius $r \leq 0.015$. At this critical r value, we have on average $N_{\text{subjets}}/N_{\text{hadrons}} \approx 0.75$. This observation is corroborated by the ROC curve in the lower panel, which shows that there is no significant performance loss in the JFN ($r = 0.01$) as compared to the PFN. This demonstrates that there is little-to-no information encoded in the very soft/collinear emissions relevant for discriminating q vs. g jets, and suggests that IRC safe inputs are sufficient for the purpose of q vs. g classification. In section 6, we will further discuss the physical interpretation of this critical r value.

Fig. 4 (right) shows the analogous JFN results for QCD vs. Z jet classification using inclusive subjet clustering. We observe again that the JFNs smoothly converge to the result of the PFN. Different than for quark vs. gluon jet tagging, we can now choose a significantly

larger subjet radius $r \sim 0.2$ without compromising the performance of the classifier. This is related to the fact that in this case, the boosted Z-boson decay products generally lead to a two-pronged jet substructure, whereas QCD (quark and gluon) jets exhibit a single-pronged jet substructure (see Fig. 3 as well as Refs. [] for different observables and perturbative calculations that characterize the radiation patterns of QCD and boosted Z jets). In general, machine-learned classifiers can make use of more information than the one- vs. two-pronged structure inside these jets; for $r \sim 0.2$, a significant fraction of the hadrons inside the jets are clustered into subjets, $N_{\text{subjets}}/N_{\text{hadrons}} \approx 0.2$. However, due to the observed saturation up to $r \sim 0.2$, we conclude that the information contained in soft and collinear emissions is significantly less relevant for this classification task compared to quark vs. gluon jet tagging. This is due to the physical scales that are relevant for the different jet classification tasks, which we will explore in more detail in section 6.

For both quark vs. gluon jet classification and QCD vs. Z jet classification, we have shown that for a range of subjet radii r, the JFN exhibits no significant difference in performance compared to the PFN. That is, the JFN classifier here is "gapless" in the sense that we smoothly approximate the PFN performance (for finite values of r). The clustering of soft and collinear emissions into subjets does not affect the performance as long as r is sufficiently small. This is in contrast to previous studies based on observables such as N-subjettiness variables or EFPs, which exhibit a small but persistent performance gap to PFNs [8, 16, 51]. The JFN provides the first example of a classifier with IRC-safe inputs that achieves equivalent performance to the IRC-unsafe PFNs. Our results are consistent with the intuitive expectation that very soft particles are uncorrelated with the hard process and are thus irrelevant to typical classification tasks in high-energy physics. The main question of our paper has thus been answered by these observations. At least for the two classification tasks considered here, we have found that IRC-safe information is sufficient to close the gap to IRC-unsafe classifiers. This was achieved by using the machine learning architecture and input type (momentum, position information) for both cases and by including subjet reclustering as a preprocessing step for the IRC-safe classifier. We note that our conclusions come with the following caveat. While we are going to identify relevant physical scales with the performance of the classifiers, it is possible that future advances in machine learning lead to more powerful algorithms that may require us to reduce the subjet radius r to match the performance of IRC-unsafe classifiers.

6 Learning physical scales

As discussed in the previous section, the performance of the JFNs matches that of the PFNs for finite values of the subjet radius r. In this section, we quantify in more detail the onset of the drop in performance when the subjet radius crosses certain physical scales. In this section we focus only on inclusive instead of exclusive subjet reconstruction since we are primarily interested in the physical scale associated with a fixed value of subjet radius r. Instead of considering the ROC and AUC curves shown in Figs. 4, we are now going to analyze the AUC for the two jet classification tasks for three different values of the

transverse momentum. This will allow us to identify the physical scales where the JFNs match the PFN performance.

The AUC for quark vs. gluon and QCD vs. Z jets classification are shown in Fig. 5 as a function of the subjet radius r. The result for the PFNs is shown as the left-most data point in both panels, which is obtained from the JFNs in the limit $r \to 0$. For the quark vs. gluon jet classification we observe that within the shown errors, the performance for quark vs. gluon jet tagging only starts to drop for subjet radii around $r \sim 0.01,\ 0.015,\ 0.025$ for the three different p_T ranges.

As we raise the transverse momentum p_T of the jet, this leads to a more boosted configuration and as such the angular separation between the hadrons is reduced. For very small values of r the samples with higher p_T lead to slightly better classification, because they contain more hadrons and the machine has more information to exploit (Check this claim). **(one way to test this is to study a fixed pT range, but separate the samples into particle multiplicity bins. –ajl)** As we raise the subjet radius the samples with a higher transverse momentum experience the AUC drop at smaller values of r.

On the other, for QCD vs. Z jets the AUC remains constant until $r \sim 0.1 - 0.2$, which is a factor of 10 larger. As already hinted, this difference arises due to the different relevant physical scales that we are sensitive to in the two cases. QCD jets do not have an intrinsic physical scale besides eventually the hadronization scale ~ 1 GeV. Instead, jets that contain the decay products of the boosted Z boson are sensitive to the Z mass. As long as the subjet radius is sufficiently small such that the decay products are reconstructed as separate subjets, instead of being merged into a single subjet, the classification performance remains unchanged compared to the PFN result. (A good place to start a discussion about the importance of IR info in classification...?)

To gain a deeper understanding of the underlying physics, in Fig. 6 we show the distributions of the opening angle θ_{12} between the first two leading subjets for both QCD and Z jets for different values of the subjet radius r and for different ranges of the jets' transverse momentum p_T. Here θ_{12} corresponds to the geometric distance in the η-ϕ plane without rescaling the distance with the jet radius R. We consider three representative values of the subjet radius that are above, below and near the transition region for the $p_T = [500, 550]$ GeV sample. The corresponding Feynman diagrams for the leading-order $Z \to q\bar{q}$ decay process and the subjet cones they get clustered in are shown in Fig. 7.

In the case of a sufficiently large subjet radius $\theta_{12} \leq r$, the $Z \to q\bar{q}$ decay products are clustered into a single subjet. In this case, the θ_{12} distributions of QCD and Z jets peak around the same value. The classifier performs poorly since the relevant physics that distinguishes the two jet samples is masked by the subjet clustering. As the subjet radius is lowered, the $Z \to q\bar{q}$ decay products are identified as separate subjets and the JFN can identify the relevant information to correctly classify QCD and Z jets. For the QCD jets, the θ_{12} distribution peaks at the smallest kinematically possible value for θ which is set by the subjet radius r. Instead, for Z jets, we observe a two-peak structure, where the dominant peak of the distribution is around $\theta_{12} \sim 0.35$ for $p_T \sim 500\ GeV$ and $\theta_{12} \sim 0.55$

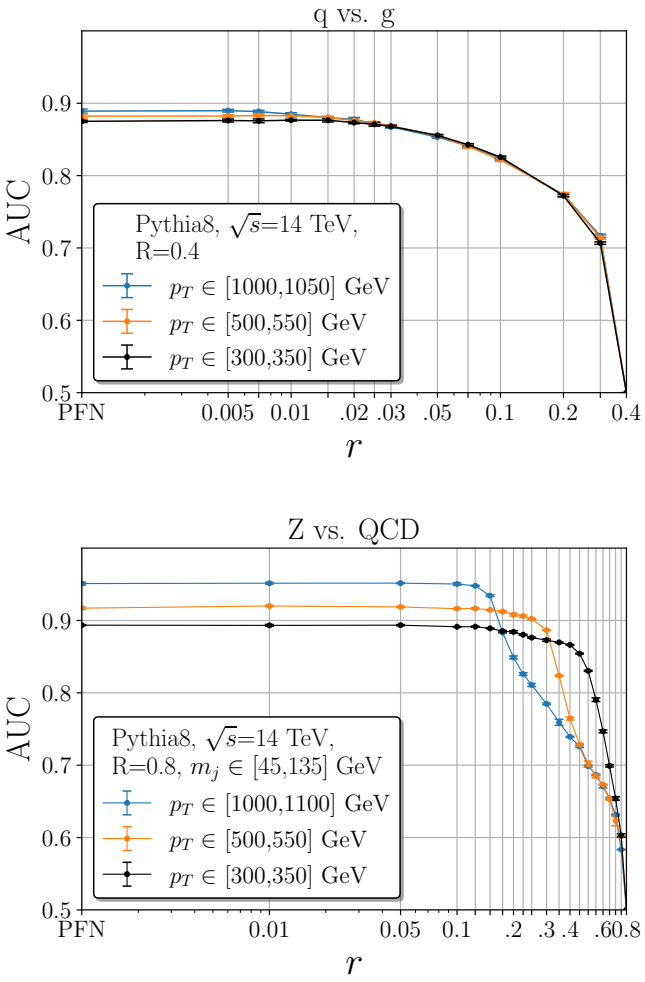

Figure 5: AUC for quark vs. gluon jets (top) and Z vs. QCD jets (bottom) for three different p_T of the jets as a function of the subjet radius r. In all six cases we used 250k jets for training, validation and testing. For q *vs.* g classification we used 250k jets for all three classifiers while for Z vs. QCD we used 500k jets. As we raise the p_T of the jet the onset of the performance drop is pushed to lower values of the subjet radius. The differences between the different curves are very small, if we put these plots side by side they are not very clear.**(maybe make the ordinate logarithmic in 1-AUC? –ajl)**

for $p_T \sim 300\ GeV$. The location of this peak is determined by M_Z and the jet's transverse momentum p_T. As long as the subjet radius is sufficiently small to resolve the Z-boson decay products lowering the subjet radius further does not increase the performance of the JFN.

Quark vs gluon (order & scaling):

$$p_T r \sim 5 \text{ GeV}\,. \tag{6.1}$$

++ comment on AUC merging for intermediate to large r values, which indicates indepen-

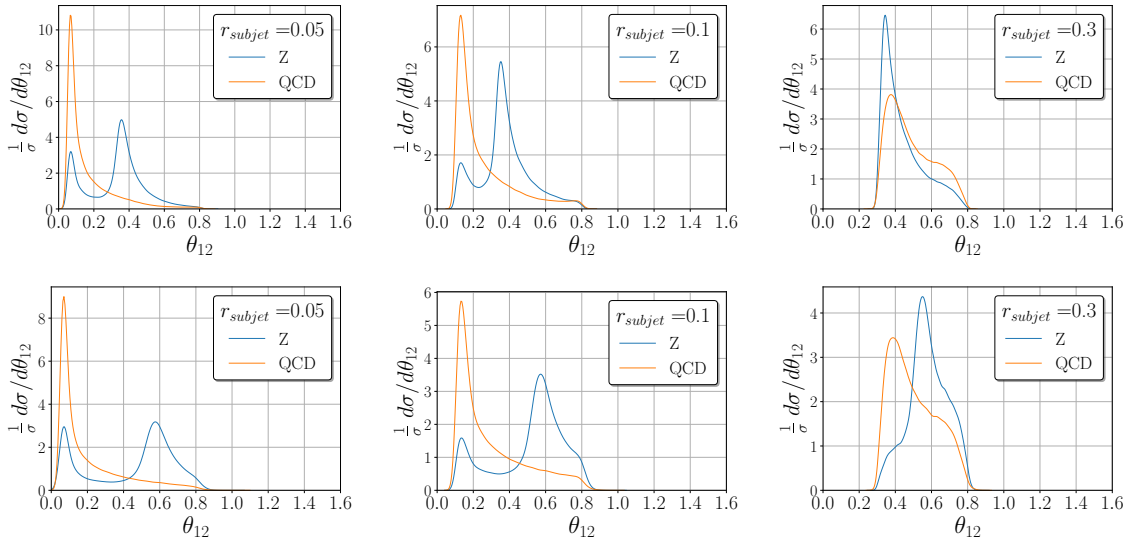

Figure 6: The distributions of the opening angle θ_{12} between the leading and first subleading subjet for both QCD and Z jets. We show the results for three different values of the subjet radius r. Upper row: Pythia distributions for $p_T = [500, 550]$ GeV. Lower row: Pythia distribution for $p_T = [300, 350]$ GeV.

dence of the resolution scale (both q vs g and QCD vs Z).

The position of the peak for the angle θ_{12} of the two leading subjets for a Z jet matches the expectation value from the Matrix Element (ME) level prediction. In particular, assuming that the decaying Z boson is on shell and that the energy is evenly split between the $q\bar{q}$ pair, then the angle between the two decaying products is fixed by the Z mass M_Z and the transverse momentum of the hard Z p_T:

$$p_T r \sim p_T \theta_{12} \sim 2M_Z \tag{6.2}$$

On the other hand, both quarks and gluons will emit QCD radiation already at the ME level. As such, the only relevant physical scale is eventually the hadronization scale. Although eq. (6.2) can not directly be applied to the case of QCD radiation, because of the $1/z$ factor in the QCD DGLAP equations [52], if we substitute $\theta_{12} \sim r_{critical} \sim 0.015$ and $p_T = 500$ GeV it correctly predicts that the critical scale r_{crit} at which the performance drop begins is associated with the hadronization scale.

Another way of illustrating the importance of resolving the two leading subjets where each subjet is matched to the Z boson's decay products produced in the hard scattering is to create a deep sets classifier that only uses the two leading (inclusive) subjets and slowly change the subjets' radius. In Fig. (8) we plot the AUC for three classifiers, the JFN, a classifier trained on the two leading inclusive subjets and a classifier trained on the three leading subjets. All classifiers are trained on the $p_T = [500, 550]$ GeV sample.

As we raise r there is a steep performance increase for N=2, 3 curves while the JFN remains in a plateau. The maximal performance for both N = 2, 3 classifiers is achieved for

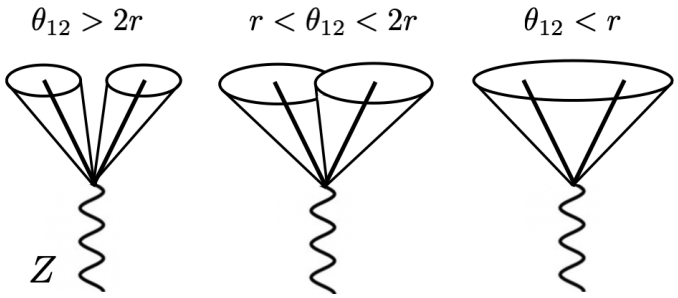

Figure 7: Reclustering of the Z-boson decay products into subjets with different radii.

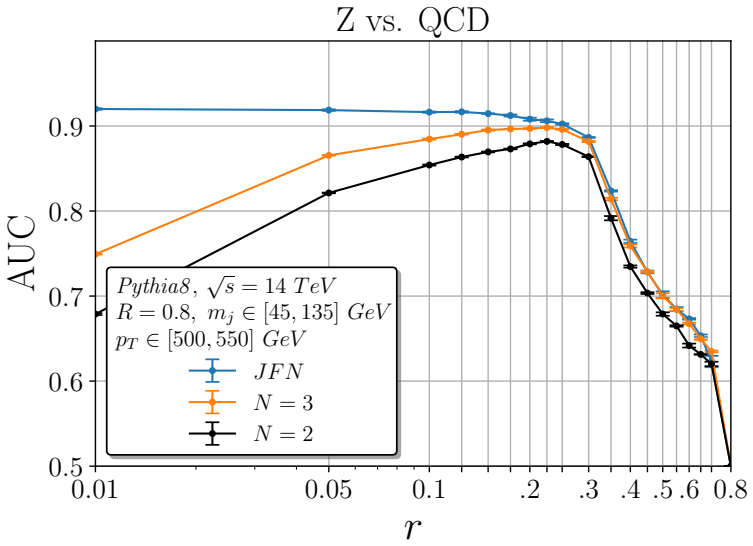

Figure 8: AUC of the JFN for QCD vs. Z jets trained on the full information (inclusive subjets) compared to JFNs trained only on the 2 or 3 leading subjets.

$r \sim 0.225 - 0.25$. As evident from Fig. (6) this is also the scale at which we start clustering the two leading decay products to the same subjet. Once we've increased r to values higher than that, all classifiers experience the same very steep drop.

Analogous classification tasks where physical scales can likely be identified are light QCD vs. c or b-jets [53–55], QCD vs. Higgs [56] or QCD vs. top quark jets [57]. We leave the exploration of these topics for future work.

7 Performance vs. generalizability

Machine learning-based classifiers are often deployed in experimental analyses to tag jet topologies. A typical method is to train the classifier using fully supervised learning on precise theoretical simulations and apply it to experimental data [55, 58, 59]. However, this approach introduces model dependence as simulations do not perfectly match the actual data. In this section, we will explore some of the systematic uncertainties associated with

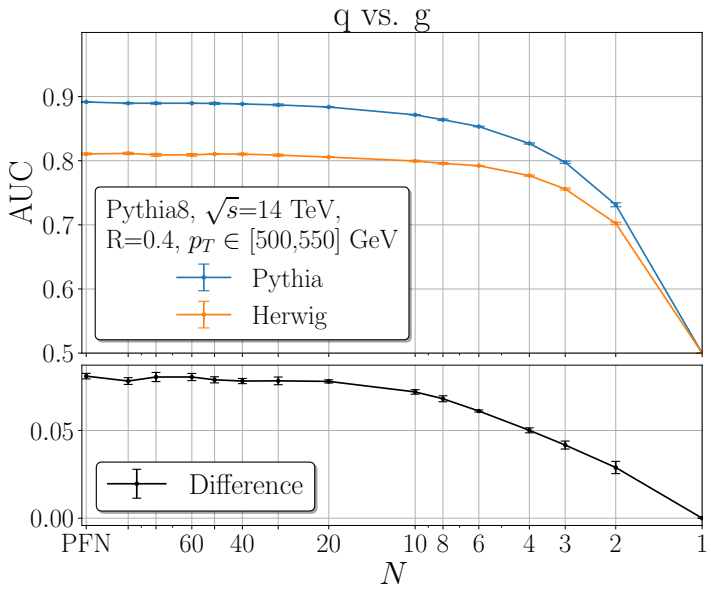

Figure 9: Classification performance for quark vs. gluon jets using JFNs and exclusive k_T clustered subjets plotted as a function of the number of subjets N. Upper panel: JFNs trained and tested on PYTHIA [42] (blue), JFNs trained on PYTHIA and tested on HERWIG [43] (orange). Lower panel: The difference in the performance of the two results.

this method. Other options that have been proposed include semi- or weakly-supervised techniques [60], as well as data-driven methods [61].

When using fully supervised learning to develop classifiers, it is crucial to ensure that the model can generalize well to the unseen experimental data. For JFNs, soft and collinear particles are clustered into subjets making them less sensitive to the modeling of IR physics. Since it is generally challenging to model the very soft physics of collider events in Monte Carlo event generators, JFNs may have an advantage compared to PFNs in terms of generalizability. On the other hand, if too many particles are clustered into few subjets, the overall performance can get worse. In order to assess whether a classifier performs well on unseen data, we train PFNs and JFNs with different parameters on PYTHIA [42] (training + validation data set) and test on HERWIG [43] simulations. Here, HERWIG can be considered as a surrogate for experimental data. We note that while the final results of both event generators are quite similar, the underlying physics of both the perturbative parton shower and the hadronization model can differ significantly. One generally expects that quark jets are quite similar in PYTHIA and HERWIG but the results for gluon jets tend to differ more significantly [62–64]. See also Ref. [65], where PYTHIA and HERWIG studies were presented using Convolutional Neural Networks (CNNs). Moreover, in Ref [66] mixed HERWIG/PYTHIA samples were combined with a Bayesian Network in order to increase model robustness.

We consider quark vs. gluon jet tagging for $p_T = [500, 550]$ GeV using exclusive k_T

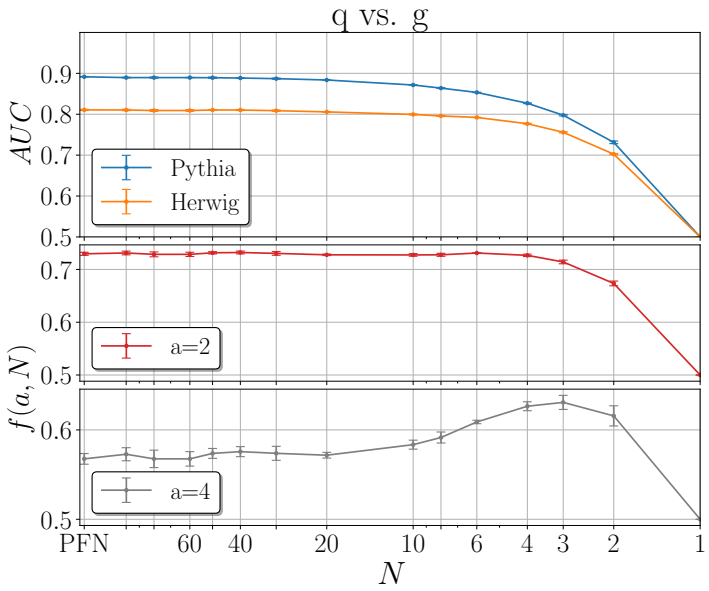

Figure 10: Upper panel: AUC of the JFNs trained on PYTHIA and tested on either PYTHIA or HERWIG plotted as a function of the number of (exclusive) subjets N, see also Fig. 9. Middle and lower panels: The objective function $f(a, N)$ defined in Eq. (7.1), where a is a weighting factor between optimal performance and generalizability.

clustering of the subjets that are taken as input to the machine learning algorithm. Fig. 9 shows the AUC as a function of the number of the subjets N. The upper panel shows the result for JFNs as a function of N trained on PYTHIA and tested on PYTHIA (blue) or HERWIG (orange). In both cases, we observe a plateau in classifier performance as the the number of (exclusive) subjets is increased. Within the shown errors, we observe that the AUC in both cases reaches its maximum value for $N \sim 30$. As expected, there is a performance gap when testing the PYTHIA-trained classifier on quark vs. gluon jets generated with HERWIG compared to testing it on PYTHIA simulations. This observation is consistent with the results of Refs. [65, 66]. However, we observe that the performance gap decreases as N decreases. To better visualize this aspect, we show in the lower panel the difference between the two AUC curves shown in the upper panel. The difference becomes smaller as N is increased indicating improved generalizability of the model. Our findings suggest that clustering particles into subjets can reduce the overall performance, but it also masks modeling uncertainties of the IR physics leading to more robust classifiers. Interestingly, we find that the difference between PYTHIA and HERWIG does not decrease for small N for QCD vs. Z jet classification (not shown).

We expect that the generalizability or robustness of machine learning-based classifiers will be useful for certain experimental applications where the trade-off between performance and generalizability needs to be considered. To illustrate this aspect, we introduce the

objective function $f(a, N)$, defined as:

$$f(a, N) = \text{AUC}_{\text{Pythia}}(N) - a \cdot (\text{AUC}_{\text{Pythia}}(N) - \text{AUC}_{\text{Herwig}}(N)) \ , \tag{7.1}$$

where N is the number of exclusive subjects. Here the performance and generalizability are combined additively and a weighting factor $a > 0$ is introduced that allows us to increase/decrease the relevance of the two metrics. An optimization problem to find the optimal balance between performance (first term in Eq. (7.1)) and generalizability (second term $\sim a$ in Eq. (7.1)) can now be formulated as follows: For a given choice of the weighting factor a, find the maximal value of the objective function $f(a, N)$. The optimal number of exclusive subjets is then given by $N_{\text{opt}} = \arg\max_N f(a, N)$. We plot $f(a, N)$ for two different values of a in Fig. 10 (middle and lower panels). We observe that as a is increased (the generalizability is weighted higher), the objective function peaks at an intermediate value of N. For example, for $a = 4$ we find $N_{\text{opt}} = 3$. While our objective function is constructed for illustration purposes, this result indicates that for certain experimental analyses that employ machine learning-based classifiers, it can be advantageous to use JFNs with a finite number of subjets to achieve the desired goals.

8 Conclusions

The classification of jets at collider experiments is relevant for a wide range of tasks in high-energy particle and nuclear physics. Over the past years, machine learning-based classifiers have been developed that can achieve impressive tagging performance. While machine learning generally outperforms traditional methods by efficiently making use of the full information content, it is often unclear where the performance difference is coming from. In particular, it had been unclear if infrared-collinear (IRC) safe classifiers can match the performance of IRC-unsafe classifiers. IRC safety is motivated by theoretical considerations ensuring that observables are tractable in perturbative QCD. In addition, it is expected that the very soft physics is uncorrelated to the hard partonic process making it unlikely to be the reason of the the performance gap that has been observed between IRC-unsafe machine learning results and traditional IRC-safe observabales. In order to address these questions, we introduced in this work a new family of classifiers, the Jet Flow Networks (JFNs). Here, particles inside a jet are first clustered into subjets and their position and momentum are taken as input to a permutation-invariant neural network (deep set). The clustering of subjets allows us to control the sensitivity to soft and collinear emissions making the input to the classifier IRC safe. As the subjet radius vanishes, we recover the IRC-unsafe Particle Flow Networks (PFNs). We investigated both inclusive and exclusive subjet clustering, which can lead to important differences depending on the application. As representative examples, we considered two classification tasks: quark vs. gluon and QCD vs. Z jet tagging. Interestingly, we observed that the JFN performance matches the IRC-unsafe PFNs for finite values of the subjet radius. This makes JFNs the first IRC-safe classifier without a performance gap to their IRC-unsafe counterpart. This observation answered the main question we aimed to address in this work and indeed IRC-safe information is sufficient for the jet classification tasks considered here. As the subjet radius is increased,

the performance of the JFNs remains unchanged (and in agreement with the PFNs) until physical thresholds are crossed. For example, for quark vs. gluon jets this threshold is set by the hadronization scale, whereas for QCD vs. Z jets it is determined by the kinematics of the Z-boson decay products. In addition, we found that JFNs may offer a decreased model dependence for certain classification tasks with only a modest tradeoff in performance. This observation may lead to interesting applications of JFNs in collider phenomenology.

Our results shed new light onto the information that machines learn in high-energy physics applications. As more powerful algorithms will be developed it will be interesting to revisit the question about the potential gap between IRC-safe and IRC-unsafe classifiers. In addition, our work represents an important step toward increasing the interpretability of machine learning methods in high-energy physics.

Acknowledgments

We thank Giacinto Piacquadio for helpful discussions. JM, MP are supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under the contract DE-AC02-05CH11231. FR was supported by the Simons Foundation under the Simons Bridge program for Postdoctoral Fellowships at SCGP and YITP award number 815892; the NSF, award number 1915093; the DOE Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab, and Old Dominion University. AL is supported in part by the UC Southern California Hub, with funding from the UC National Laboratories division of the University of California Office of the President. This research used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, *CoRR* **abs/1706.03762** (2017) [arXiv:1706.03762].

[2] 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) 1–63, [arXiv:1709.04464].

[3] R. Kogler et al., *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019), no. 4 045003, [arXiv:1803.06991].

[4] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, vol. 958. Springer, 2019.

[5] L. Lonnblad, C. Peterson, and T. Rognvaldsson, *Finding Gluon Jets With a Neural Trigger*, *Phys. Rev. Lett.* **65** (1990) 1321–1324.

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

[7] F. A. Dreyer, G. P. Salam, and G. Soyez, *The Lund Jet Plane*, *JHEP* **12** (2018) 064, [arXiv:1807.04758].

[8] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073, [arXiv:1704.08249].

[9] 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].

[10] Y. S. Lai, J. Mulligan, M. Płoskoń, and F. Ringer, *The information content of jet quenching and machine learning assisted observable design*, arXiv:2111.14589.

[11] F. Santosa and W. W. Symes, *Linear inversion of band-limited reflection seismograms*, *SIAM Journal on Scientific and Statistical Computing* **7** (1986), no. 4 1307–1330, [https://doi.org/10.1137/0907087].

[12] K. Datta and A. J. Larkoski, *Novel Jet Observables from Machine Learning*, *JHEP* **03** (2018) 086, [arXiv:1710.01305].

[13] K. Datta, A. Larkoski, and B. Nachman, *Automating the Construction of Jet Observables with Machine Learning*, *Phys. Rev. D* **100** (2019), no. 9 095016, [arXiv:1902.07180].

[14] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [arXiv:1011.2268].

[15] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093, [arXiv:1108.2701].

[16] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121, [arXiv:1810.05165].

[17] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020), no. 5 056019, [arXiv:1902.08570].

[18] V. Mikuni and F. Canelli, *ABCNet: An attention-based method for particle tagging*, *Eur. Phys. J. Plus* **135** (2020), no. 6 463, [arXiv:2001.05311].

[19] A. Romero, D. Whiteson, M. Fenton, J. Collado, and P. Baldi, *Safety of Quark/Gluon Jet Classification*, arXiv:2103.09103.

[20] Y. Lu, A. Romero, M. J. Fenton, D. Whiteson, and P. Baldi, *Resolving Extreme Jet Substructure*, arXiv:2202.00723.

[21] A. J. Larkoski, S. Marzani, and J. Thaler, *Sudakov Safety in Perturbative QCD*, *Phys. Rev. D* **91** (2015), no. 11 111501, [arXiv:1502.01719].

[22] A. J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **09** (2013) 137, [arXiv:1307.1699].

[23] K. Fraser and M. D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093, [arXiv:1803.08066].

[24] K. Lee, J. Mulligan, M. Płoskoń, F. Ringer, and F. Yuan, *Machine learning-based jet and event classification at the Electron-Ion Collider with applications to hadron structure and spin physics*, *JHEP* **03** (2023) 085, [arXiv:2210.06450].

[25] Z.-B. Kang, A. J. Larkoski, and J. Yang, *Towards a Nonperturbative Formulation of the Jet Charge*, *Phys. Rev. Lett.* **130** (2023), no. 15 151901, [arXiv:2301.09649].

[26] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, *Deep sets*, *CoRR* **abs/1703.06114** (2017) [arXiv:1703.06114].

[27] E. Wagstaff, F. B. Fuchs, M. Engelcke, I. Posner, and M. A. Osborne, *On the limitations of representing functions on sets*, *CoRR* **abs/1901.09006** (2019) [arXiv:1901.09006].

[28] B. Bloem-Reddy and Y. Teh, *Probabilistic symmetries and invariant neural networks*, *Journal of Machine Learning Research* **21** (2020), no. 90 1–61.

[29] **ALICE** Collaboration, *Measurement of inclusive and leading subjet fragmentation in pp and Pb-Pb collisions at* $\sqrt{s_{\rm NN}}$ *= 5.02 TeV*, arXiv:2204.10270.

[30] Y. Chen et al., *Jet energy spectrum and substructure in* e^+e^- *collisions at 91.2 GeV with ALEPH Archived Data*, arXiv:2111.09914.

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

[32] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally invariant* K_t *clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187–224.

[33] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [arXiv:1111.6097].

[34] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, *XCone: N-jettiness as an Exclusive Cone Jet Algorithm*, *JHEP* **11** (2015) 072, [arXiv:1508.01516].

[35] M. Dasgupta, F. Dreyer, G. P. Salam, and G. Soyez, *Small-radius jets to all orders in QCD*, *JHEP* **04** (2015) 039, [arXiv:1411.5182].

[36] L. Dai, C. Kim, and A. K. Leibovich, *Fragmentation of a Jet with Small Radius*, *Phys. Rev. D* **94** (2016), no. 11 114023, [arXiv:1606.07411].

[37] Z.-B. Kang, F. Ringer, and W. J. Waalewijn, *The Energy Distribution of Subjets and the Jet Shape*, *JHEP* **07** (2017) 064, [arXiv:1705.05375].

[38] D. Neill, F. Ringer, and N. Sato, *Leading jets and energy loss*, *JHEP* **07** (2021) 041, [arXiv:2103.16573].

[39] W. J. Waalewijn, *Calculating the Charge of a Jet*, *Phys. Rev. D* **86** (2012) 094030, [arXiv:1209.3019].

[40] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, *Jet Charge at the LHC*, *Phys. Rev. Lett.* **110** (2013), no. 21 212001, [arXiv:1209.2421].

[41] P. Komiske, E. Metodiev, and J. Thaler, *Pythia8 quark and gluon jets for energy flow*, *Zenodo* (2019).

[42] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012].

[43] J. Bellm et al., *Herwig 7.2 release note*, *Eur. Phys. J. C* **80** (2020), no. 5 452, [arXiv:1912.06509].

[44] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002, [arXiv:1004.2489].

[45] D. Napoletano and G. Soyez, *Computing* N*-subjettiness for boosted jets*, *JHEP* **12** (2018) 031, [arXiv:1809.04602].

[46] F. Chollet et al., "Keras." https://github.com/fchollet/keras, 2015.

[47] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, and et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from tensorflow.org.

[48] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (J. Fürnkranz and T. Joachims, eds.), pp. 807–814, 2010.

[49] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *CoRR* **abs/1412.6980** (2015) [arXiv:1412.6980].

[50] D. R. Cox, *The regression analysis of binary sequences*, *Journal of the Royal Statistical Society: Series B (Methodological)* **20** (1958), no. 2 215–232.

[51] 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].

[52] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*, vol. 8. Cambridge University Press, 2, 2011.

[53] **ATLAS** Collaboration, G. Aad et al., *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019), no. 11 970, [arXiv:1907.05120].

[54] **ATLAS** Collaboration, G. Aad et al., *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, *Eur. Phys. J. C* **82** (2022), no. 1 95, [arXiv:2109.10627].

[55] **ATLAS Collaboration** Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, tech. rep., CERN, Geneva, May, 2020. All figures including auxiliary figures are available in ATL-PHYS-PUB-2020-014.

[56] C. K. Khosa and S. Marzani, *Higgs boson tagging with the Lund jet plane*, *Phys. Rev. D* **104** (2021), no. 5 055043, [arXiv:2105.03989].

[57] F. A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052, [arXiv:2012.08526].

[58] **ATLAS** Collaboration, *Constituent-Based Top-Quark Tagging with the ATLAS Detector*, .

[59] **ATLAS** Collaboration, *Digluon Tagging using $\sqrt{s} = 13$ TeV pp Collisions in the ATLAS Detector*, .

[60] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, *Weakly Supervised Classification in High Energy Physics*, *JHEP* **05** (2017) 145, [arXiv:1702.00414].

[61] E. M. Metodiev and J. Thaler, *Jet Topics: Disentangling Quarks and Gluons at Colliders*, *Phys. Rev. Lett.* **120** (2018), no. 24 241602, [arXiv:1802.00008].

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

[63] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091, [arXiv:1704.03878].

[64] **ATLAS** Collaboration, G. Aad et al., *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **74** (2014), no. 8 3023, [arXiv:1405.6583].

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

[66] A. Butter, B. M. Dillon, T. Plehn, and L. Vogel, *Performance versus Resilience in Modern Quark-Gluon Tagging*, arXiv:2212.10493.