How much information is in a jet?

Kaustuv Datta and Andrew Larkoski

Physics Department, Reed College,
Portland, OR 97202, U.S.A.

E-mail: dattak@reed.edu, larkoski@reed.edu

ABSTRACT: Machine learning techniques are increasingly being applied toward data analyses at the Large Hadron Collider, especially with applications for discrimination of jets with different originating particles. Previous studies of the power of machine learning to jet physics have typically employed image recognition, natural language processing, or other algorithms that have been extensively developed in computer science. While these studies have demonstrated impressive discrimination power, often exceeding that of widely-used observables, they have been formulated in a non-constructive manner and it is not clear what additional information the machines are learning. In this paper, we study machine learning for jet physics constructively, expressing all of the information in a jet onto sets of observables that completely and minimally span N-body phase space. For concreteness, we study the application of machine learning for discrimination of boosted, hadronic decays of Z bosons from jets initiated by QCD processes. Our results demonstrate that the information in a jet that is useful for discrimination power of QCD jets from Z bosons is saturated by only considering observables that are sensitive to 4-body (8 dimensional) phase space.

Contents

1 Introduction

The problem of discrimination and identification of high energy jet-like objects observed at the Large Hadron Collider (LHC) is fundamental for both Standard Model physics and searches as the lower bound on new physics mass scales increase. Heavy particles of the Standard Model, like the W, Z, and H bosons or the top quark, can be produced with large Lorentz boosts and dominantly decay through hadrons. They will therefore appear collimated in the detector and similar to that of jets initiated by light QCD partons. The past several years have seen a huge number of observables and techniques devoted to jet identification [1–4], and many have become standard tools in the ATLAS and CMS experiments.

The list of observables for jet discrimination is a bit dizzying, and in many cases there is no organizing principle for which observables work well in what situations.[1] Motivated by the large number of variables that define the structure of a jet, several groups have recently applied machine learning methods to the problem of jet identification [9–21]. Rather than developing clever observables that identify certain physics aspects of the jets, the idea of the machine learning approach is to have a computer construct an approximation to the optimal classifier that discriminates signal from background. For example, ref. [11] interpreted the jet detected by the calorimetry as an image, with the pixels corresponding to the calorimeter cells and the "color" of the pixel corresponding to the deposited transverse momentum in

---

[1]There has been some effort in the past to identify and quantify (over)complete bases of jet observables [5–8].

the cell. These techniques have outperformed standard jet discrimination observables and show that there is additional information in jets to exploit.

However, this comes with a significant cost. Machine learning methods applied to jet physics typically have hundreds of input variables with thousands of correlations between them. Thus, in one sense this problem seems ideally suited for machine learning, but it also lacks the immediate physical interpretation and intuition that individual observables have. Previous studies have shown that the computer is learning information about what discriminates jets of different origins, but it has not been clearly demonstrated what information standard observables are missing. Along these same lines, the improvement of discrimination performance of machine learning over standard observables is relatively small, suggesting that standard observables capture the vast majority of useful information in jets.
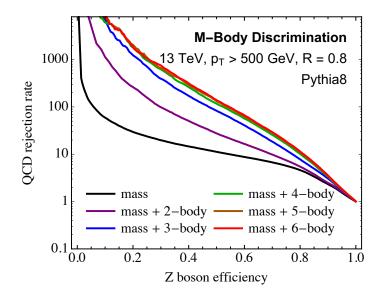
In this paper, we approach machine learning for jet discrimination from a different perspective. We construct an observable basis that completely and minimally spans the phase space for the substructure of a jet.[2] For a jet with M particles, the phase space is $3M - 4$ dimensional, and so we identify $3M - 4$ infrared and collinear (IRC) safe jet substructure observables that span the phase space.[3] These basis observables are then passed to a machine learning algorithm for identification of relevant discrimination information.[4] A general jet will have an arbitrary number of particles in it, and so we will observe how the discrimination power depends on the dimension of phase space that we assume. That is, we will assume that the jet has 2 particles, 3 particles, 4 particles, etc., as defined by the set of basis observables and observe how the discrimination power improves. This method is constructive in the following sense. With some number of assumed particles in the jet, the discrimination power will saturate, which then immediately tells us what reduced set of observables are necessary to effectively extract all information that is useful for discrimination. This approach has the additional advantage that the identified observables can be calculated theoretically from first principles, without relying on parton shower modeling.

As it is a widely-studied problem in jet substructure, we will apply this approach to the discrimination of boosted, hadronically decaying Z bosons from jets initiated by light quarks or gluons. The results of our study are shown in figure 1. Here, we plot the simulated signal (Z boson) efficiency versus the background (QCD jet) rejection rate as determined by a deep neural network, for observables that are sensitive to 2-, 3-, 4-, 5- and 6-body phase space. To identify the phase space variables, we choose to measure the jet mass and the N-subjettiness observables [24–26], but this choice is not special. This plot demonstrates that observables sensitive to 4-body phase space saturate the discrimination power. 4-body phase space is only 8 dimensional, suggesting that very few observables

---

[2]By "span" we do not mean in the vector space sense. Rather, the measurement of the basis of observables defines a system of equations that can be inverted to uniquely determine the phase space variables.

[3]Note that this will completely define the phase space of the jet substructure; that is the relative configuration of emissions in the jet. It will not identify the total jet (p_T, η, ϕ). This may be useful information, but is explicitly sensitive to global event properties which is beyond the scope of this paper. We thank Ben Nachman for emphasizing this point.

[4]Because we input a finite number of IRC safe observables to the machine, its output classifier will in general be Sudakov safe [22, 23].

Figure 1. *Z* boson jet efficiency vs. QCD jet rejection rate plot as generated by the deep neural network. Details of the event simulation, jet finding, and machine learning are described in section 3. The different curves correspond to the mass plus collections of observables that uniquely define *M*-body phase space. Discrimination power is seen to saturate when 4-body phase space is resolved.

are necessary to identify all interesting structure of these jets. We anticipate that this approach can be applied to other discrimination problems in jet substructure, as well, and greatly reduce the dimensionality of the variable space that is being studied.

The outline of this paper is as follows. In section 2, we define the observable basis that is used to identify all variables of *M*-body phase space. As mentioned above, we choose to use the *N*-subjettiness observables. In this section, we also prove that the set of observables is complete and minimal. In section 3, we discuss our event simulation and machine learning implementation. We present the results of our study, and compare discrimination power from the *M*-body phase space observables to standard observables as a benchmark. We conclude in section 4. Additional details are in the appendices.

2   Observable basis

In this section, we specify the basis of IRC safe observables that we use to identify structure in the jet. For simplicity, we will exclusively use the *N*-subjettiness observables [24–26], however this choice is not special. One could equivalently use the originally-defined *N*-point energy correlation functions [27], or their generalization to different angular dependence [28]. Our choice of using the *N*-subjettiness observbles in this analysis is mostly practical: the evaluation time for the *N*-subjettiness observables is significantly less than for the energy correlation functions. We also emphasize that the particular choice of observables below is to just ensure that they actually span the phase space for emissions in a jet. There may be a more optimal choice of a basis of observables, but optimization of the basis is beyond this paper.

The N-subjettiness observable $\tau_N^{(\beta)}$ is a measure of the radiation about N axes in the jet, specified by an angular exponent $\beta > 0$:

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \text{Jet}} p_{Ti} \min\left\{ R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta} \right\} . \tag{2.1}$$

In this expression, p_{TJ} is the transverse momentum of the jet of interest, p_{Ti} is the transverse momentum of particle i in the jet, and R_{Ki}, for $K = 1, 2, \ldots, N$, is the angle in pseudorapidity and azimuth between particle i and axis K in the jet. There are numerous possible choices for the N axes in the jet; in our numerical implementation, we choose to define them according to the exclusive k_T algorithm [29, 30] with standard E-scheme recombination [31]. Note that $\tau_N^{(\beta)} = 0$ for a jet with N or fewer particles in it.

To identify structure in the jet, we need to measure an appropriate number of different N-subjettiness observables. This requires an organizing principle to ensure that the basis of observables is complete and minimal. Our approach to ensuring this is to identify the set of N-subjettiness observables that can completely specify the coordinates of M-body phase space. Ensuring that the set is minimal is then straightforward: as M-body phase space is $3M - 4$ dimensional, we only measure $3M - 4$ N-subjettiness observables. A jet also has an overall energy scale. To ensure sensitivity to this energy scale, we will also measure the jet mass, m_J.

We will describe how to do this for low dimensional phase space, and then generalize to arbitrary M-body phase space. We will work in the limit where the jet is narrow and so all particles in the jet can be considered as relatively collinear. This simplifies the expressions for the values of the N-subjettiness observables to illustrate their content, but does not affect their ability to span the phase space variables.

- **2-body phase space.** 2-body phase space is $3 \cdot 2 - 4 = 2$ dimensional. For a jet with two particles, the phase space can be completely specified by the transverse momentum fraction z of one of the particles:

$$z = \frac{p_{T1}}{p_{TJ}}, \qquad 1 - z = \frac{p_{T2}}{p_{TJ}}, \tag{2.2}$$
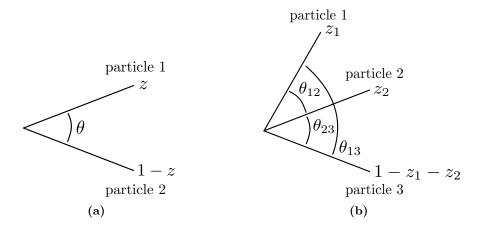
and the splitting angle θ between the particles. This configuration is shown in figure 2a. To uniquely identify the z and θ of this jet, we can measure two 1-subjettiness observables, defined by different angular exponents $\alpha \neq \beta$. For concreteness, we will measure $\tau_1^{(1)}$ and $\tau_1^{(2)}$.

To determine the measured values of the 1-subjettiness observables, we need to determine the angle between the individual particles of the jet and the axis. Because E-scheme recombination conserves momentum, the angles between the particles 1 and 2 and the axis are:

$$\theta_1 = (1 - z)\theta, \qquad \theta_2 = z\theta. \tag{2.3}$$

It then follows that the values of the 1-subjettiness observables are:

$$\tau_1^{(1)} = 2z(1 - z)\theta, \qquad \tau_1^{(2)} = z(1 - z)\theta^2. \tag{2.4}$$

Figure 2. Illustration of the momentum fraction and pairwise angle variables that describe 2-body (right) and 3-body (left) phase space.

These expressions can be inverted to find z and θ individually:

$$z(1-z) = \frac{\left(\tau_1^{(1)}\right)^2}{4\tau_1^{(2)}}, \qquad \theta = \frac{2\tau_1^{(2)}}{\tau_1^{(1)}}. \tag{2.5}$$

Note the symmetry for $z \leftrightarrow 1-z$: this is to be expected because we have not assumed an ordering of the transverse momenta of particles 1 and 2.

- **3-body phase space.** 3-body phase space is $3 \cdot 3 - 4 = 5$ dimensional, and so to completely determine the configuration of a jet with three particles, we need to measure 5 N-subjettiness observables. The 5 phase space variables can be defined to be the 3 pairwise angles between the particles i and j in the jet: θ_{12}, θ_{13}, and θ_{23}, and two of the transverse momentum fractions, say, z_1 and z_2. We define the momentum fractions as:

$$z_1 = \frac{p_{T1}}{p_{TJ}}, \qquad z_2 = \frac{p_{T2}}{p_{TJ}}, \qquad 1 - z_1 - z_2 = \frac{p_{T3}}{p_{TJ}}. \tag{2.6}$$

This configuration is shown in figure 2b. To determine the phase space variables, we will measure a collection of 1- and 2-subjettiness observables.

Our choice for which collection of 1- and 2-subjettiness observables is the following. We will measure three 1-subjettiness observables $\tau_1^{(0.5)}$, $\tau_1^{(1)}$, and $\tau_1^{(2)}$ and two 2-subjettiness observables $\tau_2^{(1)}$ and $\tau_2^{(2)}$. To motivate this collection of observables, note that one of the axes for measuring 2-subjettiness necessarily lies along the direction of a particle. Therefore, measuring 2-subjettiness is only sensitive to one relative energy fraction and one angle between pairs of particles, as illustrated explicitly in the 2-body case in eq. (2.4). Because 2-subjettiness is only sensitive to two phase space variables, we only measure two 2-subjettiness observables.

The axis for the 1-subjettiness observables, however, is necessarily displaced from the direction of any particle in the jet.[5] This is because the E-scheme recombination conserves momentum, and so this axis can only degenerate to the direction of a particle in the jet if another particle has 0 energy or is exactly collinear to another particle. Therefore, this collection of 5 N-subjettiness observables will generically span the full 3-body phase space. In appendix A, we present the explicit expressions for the 1- and 2-subjettiness observables in terms of the phase space coordinates.

- **M-body phase space.** For M-body phase space, we can define the coordinates of that phase space by $M - 1$ transverse momentum fractions z_i, for $i = 1, \ldots, M - 1$, and $2M - 3$ pairwise angles θ_{ij} between particles i and j. The remaining

$$\binom{M}{2} - (2M - 3) = \frac{1}{2}(M - 2)(M - 3),$$

pairwise angles are then uniquely determined by the geometry of points in a plane.[6] To determine all of these phase space variables, we extend the set of N-subjettinesses that were measured in the 2- and 3-body case. In this case, the $3M - 4$ observables we measure are:

$$\left\{ \tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(1)}, \tau_{M-1}^{(2)} \right\} . \qquad (2.8)$$

Note that there are $3(M-2) + 2 = 3M - 4$ observables, and these will span the space of phase space variables for generic momenta configurations, when all particles have non-zero energy and are a finite angle from one another.

As we observed in the 3-body phase space case, for a collection of M particles, all but one of the axes for the measurement of $(M - 1)$-subjettiness lies along the direction of a particle. Therefore, we only measure two $(M - 1)$-subjettiness observables. Stepping back another clustering as relevant for $(M - 2)$-subjettiness, there are two possibilities:

- Either $M - 3$ axes lie along the direction of $M - 3$ particles in the jet, and the three remaining particles are all clustered around the last axis. Then, the measurement of $(M - 2)$-subjettiness is sensitive to the phase space configuration of 3 particles in the jet. By measuring three $(M - 2)$-subjettinesses and two

---

[5]This is an important point, and the reason why we use E-scheme recombination as opposed to winner-take-all (WTA) recombination [32–34], for example, to define the N-subjettiness axes. Because the axes defined by the WTA scheme necessarily lie along the direction of particles, there are non-degenerate configurations of particles for which measuring 5 N-subjettiness observables do not span the full 3-body phase space.

[6]The proof of this is an application of the Euler Characteristic formula:

$$V - E + F = 2 . \qquad (2.7)$$

The number of vertices V is just the number of particles in the jet, M. The number of faces F is equal to the number of triangles that tesselate the plane, with vertices located at the particles. This is $F = M - 1$, as we include the face outside the region where the points are located. It then follows that the number of edges E, that is, the number of pairwise angles necessary to uniquely specify their distribution, is $E = 2M - 3$.

($M-1$)-subjettinesses, this then completely specifies the phase space configuration of those three particles.

– The other possibility is that $M-4$ axes lie along particles in the jet, while there are two particles clustered around each of the two remaining axes. About each axis, you are sensitive to the phase space configuration of two particles, which corresponds to a total of 4 phase space variables. Additionally, you are sensitive to the relative contribution of the two pairs of particles to the total ($M-2$)-subjettiness value. This configuration therefore is described by 5 phase space variables, and can be completely specified by the measurement of three ($M-2$)-subjettinesses and two ($M-1$)-subjettinesses.

This argument can be continued at further stages in the declustering. Each time an axis is removed, three new phase space variables are introduced. These can be completely specified by the measurement of three additional N-subjettiness observables. This then proves that the collection of N-subjettiness observables given above uniquely determines M-body phase space.

In the next section, we will study the information contained in this basis and use it to identify the features that are exploited in the discrimination of hadronically decaying Z boson jets from QCD jets.

3 Deep learning implementation

In this section, we describe our event simulation and implementation of machine learning to the N-subjettiness basis of observables introduced in the previous section. We generate $pp \to Z+$ jet and $pp \to ZZ$ events at the 13 TeV LHC with MadGraph5 v2.5.4 [35]. The Z boson in $pp \to Z+$ jet events is decayed to neutrinos, while one Z boson in $pp \to ZZ$ events is decayed to neutrinos, while the other is decayed to quarks. These tree-level events are then showered in Pythia v8.223 [36, 37] with default settings. In appendix B, we will show results showered with Herwig v7.0.4 [38, 39], however with one-tenth the number of events as the Pythia samples. Ignoring the neutrinos in the showered and hadronized events, we use FastJet v3.2.1 [40, 41] to cluster the jets. On the clustered anti-k_T [42] jets with radius $R = 0.8$ and minimum p_T of 500 GeV, we then measure the basis of N-subjettiness observables using the code provided in FastJet contrib v1.026. We emphasize that observables are measured on the particles as a proof of concept; we do not apply any detector simulation.

The precise set of observables we measure on the jet that we use for discrimination are the following. We measure the jet mass and the collection of N-subjettiness observables sufficient to completely determine up through 6-body phase space. That is, we measure the collection of N-subjettiness observables defined with k_T axes:

$$\left\{ \tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)} \right\} . \tag{3.1}$$

We will see that this collection of N-subjettiness observables is more than sufficient to describe all of the information useful for discrimination in the jet. Additionally, for compari-

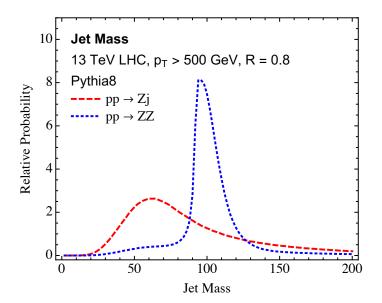| CUDA cores	Base/Boost. clock (MHz)	Memory size (GB)	Memory clock (Gbps)	Interface width	Memory Bandwidth (GB/s)
3072	1000/1075	12	7.0	384-bit	336.5

Table 1. Manufacturer specifications of the GTX TitanX.

son, we will measure a collection of standard observables that have been defined for discrimination of boosted, hadronic decays of Z bosons from jets initiated by QCD. We measure the N-subjettiness ratios $\tau_{2,1}^{(1)}$ and $\tau_{2,1}^{(2)}$ with one-pass winner-take-all (WTA) axes [32–34], and (generalized) energy correlation function ratios $D_2^{(1)}$ and $D_2^{(2)}$ [43] and $N_2^{(1)}$ and $N_2^{(2)}$ [28]. The discrimination power of these observables will provide a benchmark for the information extracted in the machine learning of the collection of N-subjettiness observables.

All deep learning analysis was carried out on the NVIDIA DIGITS DevBox, with four GeForce GTX TitanX GPUs, built on the 28 nm Maxwell architecture. The specifications of the GPU are listed in table 1. Only one GPU was used during training and testing. The dataset consisted of 7,868,000 events, split evenly between Z and QCD jets, stored in the compressed HDF5 format [44]. The data was shuffled to ensure each data file had approximately a 1:1 ratio of both classes of events. No mass cuts were imposed on the events fed to networks with the expectation that they would automatically learn the optimal cuts on mass and the observable phase space. The training and validation data consisted of 6,144,000 events and 1,536,000 events respectively, while 188,000 events were set aside for predictions. All networks were trained using the highly modular Keras [45] deep learning libraries and tested using the relevant scikit-learn [46] packages. At the time of training, data from the relevant columns of N-subjettiness variables was fed to the neural networks with the aid of a custom-designed data generator, which creates an archive of pre-processed data files.

A single neural network architecture, consisted exclusively of five fully connected layers, was utilized for all analyses. The first two Dense layers consisted of 10000 and 1000 nodes, respectively, and were assigned a Dropout [47] regularization of 0.2, while next two Dense layers consisted of 100 nodes each, and were assigned a Dropout regularization of 0.1 to prevent over-fitting on training data by making each node more 'independent'. The input layer and all hidden layers utilized the ReLU activation function [48], while the output layer, consisting of a single node, used a sigmoid activation. The network was compiled with the binary cross-entropy loss minimization function, using the Adam optimization [49]. Models were trained with Keras' default EarlyStopping, with a patience threshold of 5, to negate possible over-fitting. For each set of observables, the typical number of training epochs was about 60. To further eliminate errors due to under-training or over-training of networks, the same architecture was trained 25 different times for each round of analysis. The model that trained best for a given variable basis was picked based on a metric of maximizing the area under the signal vs. background efficiency curve.

Before showing the results from the deep neural network, we first show plots of the collection of observables sensitive to two-prong structure measured on the jets. In figure 3,

Figure 3. Distribution of the mass of the jet in $pp \rightarrow Zj$ (red dashed) and hadronically-decaying Z boson in $pp \rightarrow ZZ$ (blue dotted) from the Pythia parton shower. The minimum transverse momentum is $500\,\mathrm{GeV}$, and the jets are found with the anti-k_T algorithm with radius $R = 0.8$.

we plot the mass of the signal and background jets as defined by the simulation and jet finding from earlier. Applying a mass cut around the Z boson peak, we then measure the two-prong jet observables. In figure 4, we show the distributions of the N-subjettiness and energy correlation function ratios $\tau_{2,1}^{(\beta)}$, $D_2^{(\beta)}$, and $N_2^{(\beta)}$. As was extensively studied in the original works, these plots make clear the separation power that these observables enable. When we compare these observables to the discrimination power of the M-body phase space observables, we relax the hard mass cut, and let the machine learn the optimal mass and observable cuts dynamically.

In figure 1, we plot the signal jet (Z boson) efficiency versus the background jet (QCD) rejection rate for the collection of observables that minimally span M-body phase space, along with the jet mass. The observables that are passed to the neural network to specify M-body phase space are, explicitly:

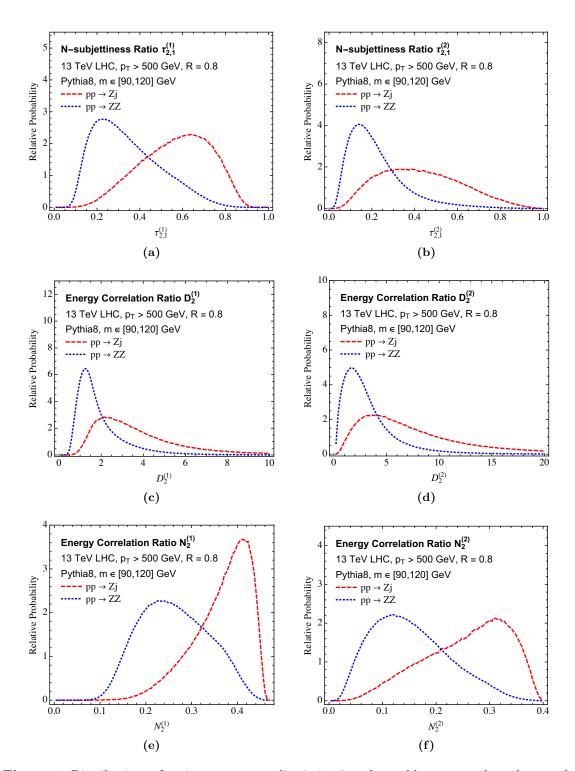2-body: $\tau_1^{(1)}, \tau_1^{(2)}$

3-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(1)}, \tau_2^{(2)}$

4-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(1)}, \tau_3^{(2)}$

5-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(1)}, \tau_4^{(2)}$

6-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)}$

Significant gains in discrimination power are observed by including observables sensitive to higher-body phase space, until enough observables to specify at least 4-body phase space are included. Including observables sensitive to 5- and 6-body phase space does not improve discrimination power, and therefore suggests that there is only an extremely limited amount of information in a jet useful for discrimination.

Figure 4. Distributions of various two-prong discrimination observables measured on the sample of jets showered with Pythia, on which a mass cut of $m \in [90, 120]$ GeV has been placed. From top to bottom are plotted signal (blue dotted) and background (red dashed) distributions of: N-subjettiness ratios $\tau_{2,1}^{(1)}$ (left) and $\tau_{2,1}^{(2)}$ (right), energy correlation function ratios $D_2^{(1)}$ (left) and $D_2^{(2)}$ (right), and $N_2^{(1)}$ (left) and $N_2^{(2)}$ (right).

To see what information is necessary to accomplish the maximal discrimination power, in figure 5 we plot the signal efficiency versus background rejection rate for the collection of N-subjettiness and energy correlation function ratios plotted earlier. For comparison, we also include the corresponding curves for the jet mass, jet mass plus 3-body phase space observables, and jet mass plus 4-body phase space observables. The discrimination power of all of these observables are comparable, and this illustrates that they appear to capture most of the information contained in the 3-body phase space observables. Then, to match the maximum discrimination power (as represented by the jet mass plus 4-body phase space curve), one just needs to augment the measurement of jet mass and an N-subjettiness or energy correlation function ratio with observables that are sensitive to some 3- and 4-body phase space information. We leave the construction of these optimal 3- and 4-body phase space observables for this purpose to future work.

As a cross check that our minimal basis of N-subjettiness observables listed above does capture the maximal amount of information useful for discrimination, in figure 6, we compare our minimal basis to an overcomplete basis of observables. Here, we measure the mass and the following collection of N-subjettiness observables on the jet:
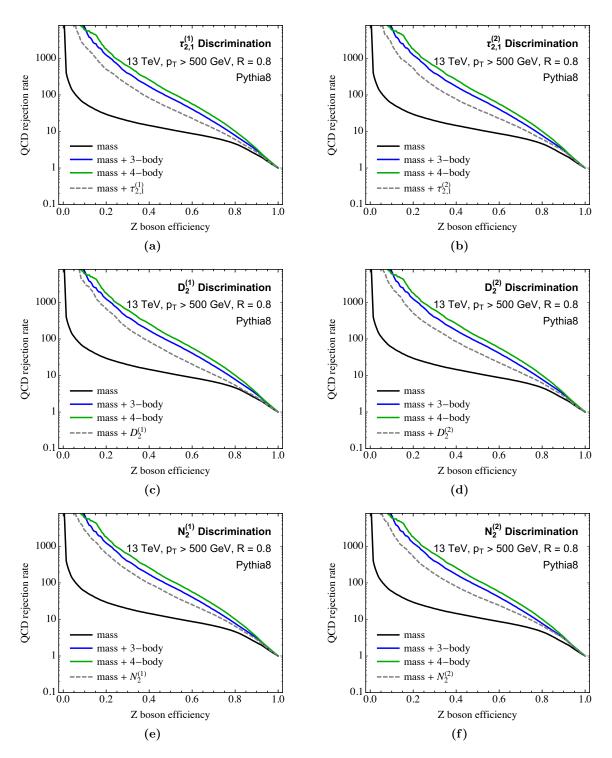
$$\left\{ \tau_1^{(0.25)}, \tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_1^{(4)}, \tau_2^{(0.25)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_2^{(4)}, \tau_3^{(0.25)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_3^{(4)}, \right. \quad (3.2)$$
$$\left. \tau_4^{(0.25)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_4^{(4)} \right\} .$$

From our arguments in section 2, this is an overcomplete basis for 5-body phase space and therefore should not contain any additional information useful for discrimination. This is illustrated in figure 6 where we plot the discrimination power of this overcomplete basis as determined by the neural network described earlier. For comparison, we also show the discrimination power of the jet mass, the jet mass plus the 3-body observable basis, and the jet mass plus the 4-body observable basis as determined by the neural network described earlier. As expected, no improvement of discrimination power is accomplished when more observables beyond the minimal set are included. The apparent slight decrease in discrimination power using the overcomplete basis is likely due to suboptimal training because of the large number of input observables.
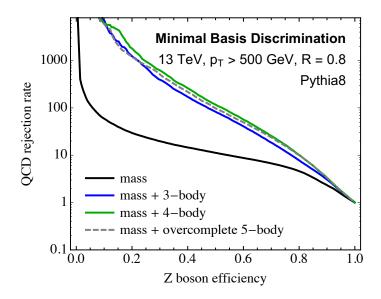
In appendix C, we present results for the signal vs. background efficiency as determined by a neural network with an additional hidden layer and the result of a boosted decision tree. These different classification networks demonstrate the same conclusion, that discrimination power saturates once enough observables are measured to resolve 4-body phase space. Additionally, these results show that the discrimination power of the overcomplete basis is just marginally better than that accomplished by the 4-body observable basis. This is consistent with our observation that 4-body phase space is essentially saturating all useful discrimination information.

4 Conclusions

Motivated by both the enormous data sets produced by the ATLAS and CMS experiments as well as their exceptional resolution, deep learning approaches to physics at the LHC are

Figure 5. Signal efficiency versus background rejection rate for jet mass plus N-subjettiness ratios $\tau_{2,1}^{(1)}$ (left) and $\tau_{2,1}^{(2)}$ (right), energy correlation function ratios $D_2^{(1)}$ (left) and $D_2^{(2)}$ (right), and $N_2^{(1)}$ (left) and $N_2^{(2)}$ (right), as determined by the neural network. For comparison, we also include the signal efficiency versus background rejection rate for jet mass, jet mass plus 3-body phase space observables, and jet mass plus 4-body phase space observables.

Figure 6. Signal efficiency versus background rejection rate for jet mass plus the overcomplete basis of observables that are sensitive to 5-body phase space described in the text, as determined by the neural network. For comparison, we also include the signal efficiency versus background rejection rate for jet mass, jet mass plus minimal 3-body phase space observables, and jet mass plus the minimal 4-body phase space observables.

seeing an increased interest. This is especially true for jet physics, where the identification of the initiating particle of a jet is of fundamental importance. Previous applications of deep learning to jet physics applied techniques from computer science (like image recognition or natural language processing) and demonstrated impressive discrimination power. While the effectiveness of these methods is exceptional, they often lack a physical interpretation and are not presented in a constructive manner. The deep neural network is definitely identifying relevant structure in the jets, but what this is or if it is just a feature of the simulated data is not identified. Other recent efforts to reduce dependence on modeling have been studied in the context of weak supervision in ref. [20].

In this paper, we have approached the problem of machine learning for jet physics in a physically clear, constructive manner. Instead of providing the machine with the energy deposits in calorimeter cells of the jet, we measure a basis of observables on the jet that completely and minimally spans M-body phase space. The effective resolution to the emissions in the jet is increased by increasing the number of observables measured on the jet. We demonstrated that the information useful for discrimination of a jet initiated by a boosted, hadronically-decaying Z boson from a jet initiated by a light QCD parton is saturated when enough observables are measured to span 4-body phase space. As 4-body phase space is only 8 dimensional, the amount of useful information in the jet is quite small. Additionally, this procedure is constructive in the sense that one can then form observables that are non-zero for a jet with four constituents to optimally discriminate signal from background. Similar constructions of observable bases for identifying particular phase space regions has been studied recently to resum non-global logarithms [50] and calculate multi-differential cross sections on jets [51, 52].

Important for our analysis is that we use an IRC safe basis of observables that span the M-body phase space, namely, the N-subjettiness observables. This is vital for constructibility, as in principle the cross section for the measurement of multiple N-subjettiness observables on a jet can be calculated in the perturbation theory of QCD.[7] It would be possible to additionally include information that is not IRC safe, for example, jet charge. Nevertheless, some non-IRC safe information is already included in this approach, like the jet constituent multiplicity. Additionally, included in the basis of M-body phase space observables are techniques like jet grooming that systematically remove radiation from the jet. This could enable a systematic study of how jet grooming methods affect the optimal discrimination observables, which has been addressed recently [28, 54].

An advantage of our approach is that the jet data is preprocessed in a useful way at the same time that the basis observables are being measured. In applications of image processing to jets, one typically has to perform a series of transformations to ensure that different jets can be compared (see the discussion in, e.g., ref. [11]). Jets must be rotated and rescaled appropriately so that (approximate) symmetries do not wash out the ability to discriminate. By instead measuring a collection of IRC safe observables like N-subjettiness on which we train, this preprocessing step is unnecessary, as the value of the observable is only sensitive to relative angles between particles and energy fractions.

From our results, it would also be interesting to study in detail the information for discrimination that is missed when using standard jet observables like N-subjettiness ratios $\tau_{2,1}^{(\beta)}$ or energy correlation function ratios $D_2^{(\beta)}$ or $N_2^{(\beta)}$. The construction and justification of these particular observables exploited properties of QCD in the soft and/or collinear limits. These observables appear to be sensitive to most of the 3-body phase space information available for discrimination of boosted, hadronically decaying Z bosons from QCD jets. Observables that are sensitive to the remaining information for discrimination could be constructed by studying in detail the differences between how the decays of Z bosons and QCD fill 4-body phase space. We anticipate that these methods can also be used for discrimination of many different types of jets, including quark versus gluon and QCD versus top quark discrimination, as well as for multi-label classification of jets. The ultimate goal of such a program would be to design an anti-QCD tagger which could identify, using only a few observables that are sensitive to a small phase space, if a jet was likely initiated by a light QCD parton. This could open the door to new classes of observables that are sensitive to exotic configurations within jets.

Acknowledgments

---

[7]Actually calculating distributions of N-subjettiness observables in practice may be a significant challenge, however [53].

A Explicit expressions for 3-body phase space

In this appendix, we present the explicit expressions for the 1- and 2-subjettiness observables measured on a jet with three particles. The configuration of particles in the jet is shown in figure 2b. We will start with the evaluation of the 2-subjettiness observables, and then the 1-subjettiness observables.

A.1 2-subjettiness

For measuring 2-subjettiness, we identify two axes defined by the exclusive k_T algorithm with E-scheme recombination. For three particles, one of the axes must necessarily lie along the direction of one particle in the jet, which we can take to be particle 3 without loss of generality. Then, only particles 1 and 2 can contribute to 2-jettiness. Call the axis about which particles 1 and 2 are clustered $\hat{A}$. Then, from figure 2b, the angle that particles 1 and 2 make with $\hat{A}$ are:

$$\theta_{1\hat{A}} = \frac{z_2}{z_1 + z_2}\theta_{12}, \qquad \theta_{2\hat{A}} = \frac{z_1}{z_1 + z_2}\theta_{12}. \tag{A.1}$$

The 2-subjettiness observables that we measure are then:

$$\tau_2^{(1)} = z_1 \cdot \frac{z_2}{z_1 + z_2}\theta_{12} + z_2 \cdot \frac{z_1}{z_1 + z_2}\theta_{12} = \frac{2z_1 z_2}{z_1 + z_2}\theta_{12}, \tag{A.2}$$

$$\tau_2^{(2)} = z_1\left(\frac{z_2}{z_1 + z_2}\theta_{12}\right)^2 + z_2 \cdot \left(\frac{z_1}{z_1 + z_2}\theta_{12}\right)^2 = \frac{z_1 z_2}{z_1 + z_2}\theta_{12}^2. \tag{A.3}$$

Therefore the values of the 2-subjettiness observables can be inverted to determine the relative momentum fraction

$$\frac{z_1 z_2}{z_1 + z_2},$$

and the pairwise angle θ_{12}.

A.2 1-subjettiness

Now, we would like to calculate the value of 1-subjettiness on this configuration of particles. This requires determining the angle between each of the three particles and their direction of net momentum. To determine these angles, we consider the distribution of particles in the jet in a plane, as displayed in figure 7. We set particle 1 at the origin $(0,0)$ of the plane, particle 2 along the horizontal axis at $(\theta_{12}, 0)$, and particle 3 at a generic point in the plane. The horizontal and vertical coordinates of particle 3 can be calculated to be:

$$\text{particle 3:}\ \left(\frac{\theta_{12}^2 + \theta_{13}^2 - \theta_{23}^3}{2\theta_{12}}, \frac{\sqrt{2\theta_{12}^2\theta_{23}^2 + 2\theta_{12}^2\theta_{13}^2 + 2\theta_{13}^2\theta_{23}^2 - \theta_{12}^4 - \theta_{23}^4 - \theta_{13}^4}}{2\theta_{12}}\right). \tag{A.4}$$

With this expression, we can determine the location of the jet axis. With E-scheme recombination, the jet axis is located at the momentum-weighted centroid of the three particles:

$$\text{jet center:}\ \left(z_2\theta_{12} + z_3\frac{\theta_{12}^2 + \theta_{13}^2 - \theta_{23}^3}{2\theta_{12}}, z_3\frac{\sqrt{2\theta_{12}^2\theta_{23}^2 + 2\theta_{12}^2\theta_{13}^2 + 2\theta_{13}^2\theta_{23}^2 - \theta_{12}^4 - \theta_{23}^4 - \theta_{13}^4}}{2\theta_{12}}\right). \tag{A.5}$$
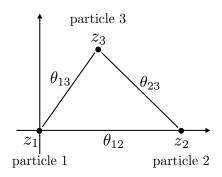
Figure 7. Configuration in a plane of a jet with three particles. Pairwise angles θ_{ij} and momentum fractions z_i of the individual particles are labeled. Momentum conservation enforces that $z_3 = 1 - z_1 - z_2$.

Here, for conciseness, we express $z_3 = 1 - z_1 - z_2$. It then follows that the angle from each particle to this jet axis $\hat{A}$ is:

$$\theta_{1\hat{A}}^2 = z_2^2\theta_{12}^2 + z_3^2\theta_{13}^2 + z_2 z_3(\theta_{12}^2 + \theta_{13}^2 - \theta_{23}^2)\,, \tag{A.6}$$

$$\theta_{2\hat{A}}^2 = z_1^2\theta_{12}^2 + z_3^2\theta_{23}^2 + z_1 z_3(\theta_{12}^2 + \theta_{23}^2 - \theta_{13}^2)\,, \tag{A.7}$$

$$\theta_{3\hat{A}}^2 = z_1^2\theta_{13}^2 + z_2^2\theta_{23}^2 + z_1 z_2(\theta_{13}^2 + \theta_{23}^2 - \theta_{12}^2)\,. \tag{A.8}$$

The values of the three 1-subjettiness observables are then:

$$\tau_1^{(0.5)} = z_1\theta_{1\hat{A}}^{0.5} + z_2\theta_{2\hat{A}}^{0.5} + z_3\theta_{3\hat{A}}^{0.5}\,, \tag{A.9}$$

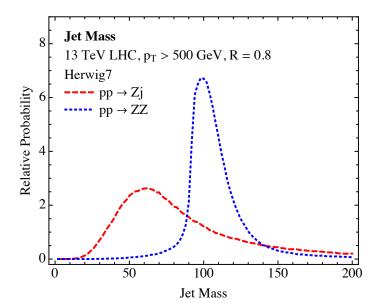$$\tau_1^{(1)} = z_1\theta_{1\hat{A}} + z_2\theta_{2\hat{A}} + z_3\theta_{3\hat{A}}\,, \tag{A.10}$$

$$\tau_1^{(2)} = z_1\theta_{1\hat{A}}^2 + z_2\theta_{2\hat{A}}^2 + z_3\theta_{3\hat{A}}^2 = z_1 z_2\theta_{12}^2 + z_1 z_3\theta_{13}^2 + z_2 z_3\theta_{23}^2\,. \tag{A.11}$$

For $\tau_1^{(2)}$, the expression simplifies significantly in terms of the momentum fractions and pairwise angles.

B Herwig results

In this appendix, we present discrimination results for jets showered in Herwig 7.0.4 [38, 39] from events generated in MadGraph. The number of events showered in Herwig is about a factor of 10 fewer than that shown in the main body of the paper with Pythia, and so the neural network training is not as efficient. Nevertheless, the conclusions drawn from this reduced Herwig sample are the same as from Pythia; namely, that observables sensitive to 4-body phase space saturate discrimination power.

On the sample of jets from $pp \to ZZ$, with one Z decaying hadronically, and $pp \to Z+$ jet, we identify the same jets and measure the same collection of N-subjettiness observables as described in the main text. These observables are then passed through the deep neural network as described in section 3, with 390,000 events each for $pp \to ZZ$ and $pp \to Z+$ jet processes. These events were divided into 684,000 used for training, 76,000 for validation, and 20,000 for testing.

Figure 8. Distribution of the mass of the jet in $pp \to Zj$ (red dashed) and hadronically-decaying Z boson in $pp \to ZZ$ (blue dotted) from the Herwig parton shower. The minimum transverse momentum is $500\,\mathrm{GeV}$, and the jets are found with the anti-k_T algorithm with radius $R = 0.8$.

In figures 8 and 9, we show validation plots on the jets showered with Herwig, to be compared with figures 3 and 4 from Pythia. The jet mass distribution in figure 8 agrees qualitatively well with the corresponding plot from Pythia; though the Herwig sample seems to lack the small shoulder of the Z boson mass distribution present in Pythia. With a cut on the jet mass around the location of the Z boson peak, we then measure the same selection of one- versus two-prong discriminant variables in the Herwig sample. Again, good qualitative agreement is seem with Pythia, though the effects of finite statistics are much more evident.

Figure 10 shows the signal efficiency vs. background rejection rate for the collections of observables that resolve M-body phase space as determined by the neural network. Just like in the Pythia samples, the discrimination power is observed to increase as more N-subjettiness observables are included. The discrimination power is observed to saturate with observables that are sensitive to 3- or 4-body phase space. This difference from when the Pythia events saturated could be due to the smaller jet sample size, though it could also be due to differences between the Pythia and Herwig parton showers. It has been observed in numerous other studies [43, 55–62] that the discrimination performance differs significantly between jets showered in Pythia versus Herwig. The exact reason for the discrepancy is beyond this paper, but the existence of a saturation point also in Herwig demonstrates that there is only a very limited amount of information in the jet for discrimination.

C   Results with other architectures

In this appendix, we show discrimination results for a neural network with one more hidden layer than the network studied in the body of the paper, as well as the output of a boosted decision tree.
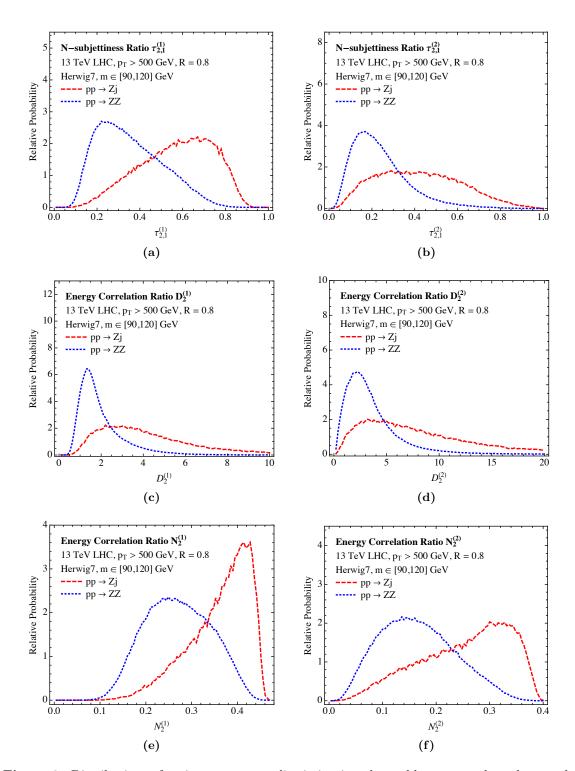
Figure 9. Distributions of various two-prong discrimination observables measured on the sample of jets showered with Herwig, on which a mass cut of $m \in [90, 120]$ GeV has been placed. From top to bottom are plotted signal (blue dotted) and background (red dashed) distributions of: N-subjettiness ratios $\tau_{2,1}^{(1)}$ (left) and $\tau_{2,1}^{(2)}$ (right), energy correlation function ratios $D_2^{(1)}$ (left) and $D_2^{(2)}$ (right), and $N_2^{(1)}$ (left) and $N_2^{(2)}$ (right).
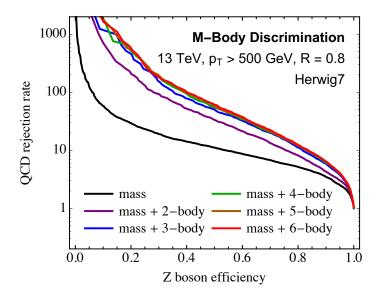
Figure 10. *Z* boson jet efficiency vs. QCD jet rejection rate plot generated by the deep neural network for jets showered in Herwig. The different curves correspond to the mass plus collections of *N*-subjettiness observables that uniquely define *M*-body phase space. Discrimination power is seen to saturate when 3- or 4-body phase space is resolved.

C.1 A deeper neural network

The neural network used in this appendix is identical to the network studied in the body of the paper, except with the addition of another layer. Immediately after the input layer, we have included an additional Dense layer of 1000 nodes, with a Dropout regularization of 0.2. The typical number of training epochs of this new neural network was about 50 for each collection of observables.

We show the discrimination performance as identified by this network in figures 11 and 12. In figure 11, we show the discrimination power as more observables are added to resolve higher-body phase space. As with the other studies in this paper, we see that the discrimination power is saturated when 4-body phase space is resolved. Additionally in figure 12, we compare the discrimination power of 3- and 4-body phase space observables to the overcomplete 5-body phase space observables described in section 3. The overcomplete basis of observables is observed to be only very slightly better than 4-body phase space basis, suggesting that essentially all useful discrimination information has been extracted.

C.2 Boosted decision tree

Because our observable basis is quite small, we can input them to a boosted decision tree to evaluate the discrimination power. We used ROOT's TMVA package [63, 64] to train and test the boosted decision trees. Each collection of phase space observables studied elsewhere in this paper were input to the boosted decision trees, and forests of 2500 trees were used. We also trained on forests of 850 trees, and observed no significant improvement in discrimination power in extending to forests of 2500 trees, suggesting that the boosted decision trees are extracting all the information that they can. The results of the boosted
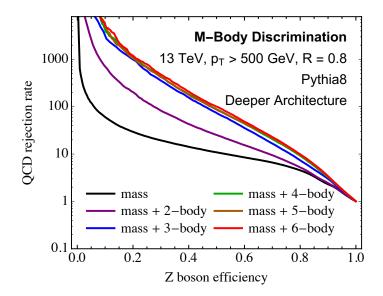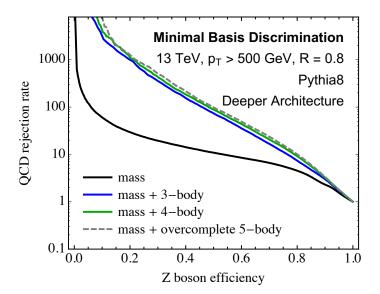
Figure 11. *Z* boson jet efficiency vs. QCD jet rejection rate plot as generated by the deeper neural network for events showered in Pythia. The different curves correspond to the mass plus collections of observables that uniquely define *M*-body phase space. Discrimination power is seen to saturate when 4-body phase space is resolved.

Figure 12. Signal efficiency versus background rejection rate for jet mass plus the overcomplete basis of observables that are sensitive to 5-body phase space described in the text, as determined by the deeper neural network. For comparison, we also include the signal efficiency versus background rejection rate for jet mass, jet mass plus minimal 3-body phase space observables, and jet mass plus the minimal 4-body phase space observables.
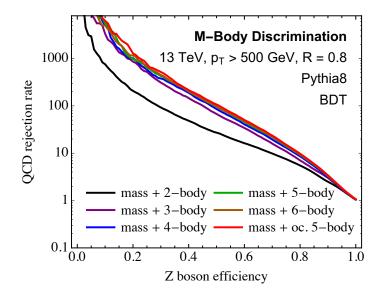
Figure 13. *Z* boson jet efficiency vs. QCD jet rejection rate plot as generated by the boosted decision tree for events showered in Pythia. The different curves correspond to the mass plus collections of observables that uniquely define *M*-body phase space. Discrimination power is seen to saturate when 4-body phase space is resolved. In this plot, we also include the overcomplete 5-body phase space collection of observables, labeled "oc. 5-body".

decision trees are shown in figure 13. These results are again consistent with what we found earlier; namely, that discrimination power is observed to saturate once 4-body phase space is resolved.

References

[1] D. Adams et al., *Towards an Understanding of the Correlations in Jet Substructure*, *Eur. Phys. J.* **C 75** (2015) 409 [arXiv:1504.00679] [INSPIRE].

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

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

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

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

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

[7] P.S. Cherzor and N.A. Sveshnikov, *Jet observables and energy momentum tensor*, in *Quantum field theory and high-energy physics. Proceedings, Workshop, QFTHEP'97, Samara, Russia, September 4–10, 1997*, pp. 402–407, (1997), `hep-ph/9710349` [INSPIRE].

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

[9] J. Cogan, M. Kagan, E. Strauss and A. Schwarztman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [`arXiv:1407.5675`] [INSPIRE].

[10] L.G. Almeida, M. Backovic, M. Cliche, S.J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086 [`arXiv:1501.05968`] [INSPIRE].

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

[12] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D 93** (2016) 094034 [`arXiv:1603.09349`] [INSPIRE].

[13] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D 94** (2016) 112002 [`arXiv:1607.08633`] [INSPIRE].

[14] J.S. Conway, R. Bhaskar, R.D. Erbacher and J. Pilot, *Identification of High-Momentum Top Quarks, Higgs Bosons and W and Z Bosons Using Boosted Event Shapes*, *Phys. Rev.* **D 94** (2016) 094027 [`arXiv:1606.06859`] [INSPIRE].

[15] J. Barnard, E.N. Dawe, M.J. Dolan and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, *Phys. Rev.* **D 95** (2017) 014018 [`arXiv:1609.00607`] [INSPIRE].

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

[17] L. de Oliveira, M. Paganini and B. Nachman, *Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis*, `arXiv:1701.05927` [INSPIRE].

[18] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006 [`arXiv:1701.08784`] [INSPIRE].

[19] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, `arXiv:1702.00748` [INSPIRE].

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

[21] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, `arXiv:1704.02124` [INSPIRE].

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

[23] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov Safety in Perturbative QCD*, *Phys. Rev.* **D 91** (2015) 111501 [`arXiv:1502.01719`] [INSPIRE].

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

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

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

[27] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [ɪɴSPIRE].

[28] I. Moult, L. Necib and J. Thaler, *New Angles on Energy Correlation Functions*, *JHEP* **12** (2016) 153 [arXiv:1609.07483] [ɪɴSPIRE].

[29] 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 [ɪɴSPIRE].

[30] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160 [hep-ph/9305266] [ɪɴSPIRE].

[31] G.C. Blazey et al., *Run II jet physics*, in *QCD and weak boson physics in Run II. Proceedings, Batavia, U.S.A., March 4-6, June 3-4, November 4-6, 1999*, pp. 47–77, (2000), hep-ex/0005012 [ɪɴSPIRE].

[32] D. Bertolini, T. Chan and J. Thaler, *Jet Observables Without Jet Algorithms*, *JHEP* **04** (2014) 013 [arXiv:1310.7584] [ɪɴSPIRE].

[33] A.J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **04** (2014) 017 [arXiv:1401.2158] [ɪɴSPIRE].

[34] A.J. Larkoski and J. Thaler, *Aspects of jets at 100 TeV*, *Phys. Rev.* **D 90** (2014) 034010 [arXiv:1406.7011] [ɪɴSPIRE].

[35] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [ɪɴSPIRE].

[36] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [ɪɴSPIRE].

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

[38] M. Bahr et al., *HERWIG++ Physics and Manual*, *Eur. Phys. J.* **C 58** (2008) 639 [arXiv:0803.0883] [ɪɴSPIRE].

[39] J. Bellm et al., *HERWIG 7.0/HERWIG++ 3.0 release note*, *Eur. Phys. J.* **C 76** (2016) 196 [arXiv:1512.01178] [ɪɴSPIRE].

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

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

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

[43] A.J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, *JHEP* **12** (2014) 009 [arXiv:1409.6298] [ɪɴSPIRE].

[44] The HDF Group, *Hierarchical Data Format, version 5*, 1997-NNNN, http://www.hdfgroup.org/HDF5/.

[45] F. Chollet, *Keras*, https://github.com/fchollet/keras, (2015).

[46] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Mach. Learn. Res.* **12** (2011) 2825.

[47] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, *J. Mach. Learn. Res.* **15** (2014) 1929.

[48] V. Nair and G.E. Hinton, *Rectified linear units improve restricted boltzmann machines.*, in *ICML*, J. Fürnkranz and T. Joachims eds., Omnipress, (2010), pp. 807–814.

[49] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, `arXiv:1412.6980`.

[50] A.J. Larkoski, I. Moult and D. Neill, *Non-Global Logarithms, Factorization and the Soft Substructure of Jets*, *JHEP* **09** (2015) 143 [`arXiv:1501.04596`] [INSPIRE].

[51] A.J. Larkoski, I. Moult and D. Neill, *Toward Multi-Differential Cross sections: Measuring Two Angularities on a Single Jet*, *JHEP* **09** (2014) 046 [`arXiv:1401.4458`] [INSPIRE].

[52] M. Procura, W.J. Waalewijn and L. Zeune, *Resummation of Double-Differential Cross sections and Fully-Unintegrated Parton Distribution Functions*, *JHEP* **02** (2015) 117 [`arXiv:1410.6483`] [INSPIRE].

[53] A.J. Larkoski and I. Moult, *The Singular Behavior of Jet Substructure Observables*, *Phys. Rev.* **D 93** (2016) 014017 [`arXiv:1510.08459`] [INSPIRE].

[54] G.P. Salam, L. Schunk and G. Soyez, *Dichroic subjettiness ratios to distinguish colour flows in boosted boson tagging*, *JHEP* **03** (2017) 022 [`arXiv:1612.03917`] [INSPIRE].

[55] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Jet Substructure*, *JHEP* **04** (2013) 090 [`arXiv:1211.7038`] [INSPIRE].

[56] ATLAS collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7\,TeV$ with the ATLAS detector*, *Eur. Phys. J.* **C 74** (2014) 3023 [`arXiv:1405.6583`] [INSPIRE].

[57] A.J. Larkoski, J. Thaler and W.J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, *JHEP* **11** (2014) 129 [`arXiv:1408.3122`] [INSPIRE].

[58] E. Izaguirre, B. Shuve and I. Yavin, *Improving Identification of Dijet Resonances at Hadron Colliders*, *Phys. Rev. Lett.* **114** (2015) 041802 [`arXiv:1407.7037`] [INSPIRE].

[59] A.J. Larkoski, I. Moult and D. Neill, *Analytic Boosted Boson Discrimination*, *JHEP* **05** (2016) 117 [`arXiv:1507.03018`] [INSPIRE].

[60] J.R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015*, (2016), `arXiv:1605.04692` [INSPIRE].

[61] ATLAS collaboration, *Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8\,TeV$ pp collisions with the ATLAS detector*, *Eur. Phys. J.* **C 76** (2016) 322 [`arXiv:1602.00988`] [INSPIRE].

[62] P. Gras et al., *Systematics of quark/gluon tagging*, `arXiv:1704.03878` [INSPIRE].

[63] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, `PoS(ACAT)040` [`physics/0703039`] [INSPIRE].

[64] P. Speckmayer, A. Hocker, J. Stelzer and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, *J. Phys. Conf. Ser.* **219** (2010) 032057 [INSPIRE].