Spectral Clustering for Jet Physics

S. Dasmahapatra*[1], H.A. Day-Hall†[2,3], B. Ford‡[2],
S. Moretti§[2,3], C.H. Shepherd-Themistocleous¶[3]

[1] *School of Electronics and Computer Science, University of Southampton,
Southampton, SO17 1BJ, United Kingdom*

[2] *School of Physics and Astronomy, University of Southampton,
Southampton, SO17 1BJ, United Kingdom*

[3] *Particle Physics Department, Rutherford Appleton Laboratory,
Chilton, Didcot, Oxon OX11 0QX, United Kingdom*

March 23, 2021

Abstract

We present a new method of jet definition as an alternative to the clustering methods, such as anti-k_T, that are based directly on kinematic data by, instead, using kinematic relations to define a spectral representation of the particles. The performance of this algorithm in analysing $gg \to H_{125\,\mathrm{GeV}} \to H_{40\,\mathrm{GeV}} H_{40\,\mathrm{GeV}} \to b\bar{b}b\bar{b}$, $gg \to H_{500\,\mathrm{GeV}} \to H_{125\,\mathrm{GeV}} H_{125\,\mathrm{GeV}} \to b\bar{b}b\bar{b}$ and $gg, q\bar{q} \to t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}jj\ell\nu_\ell$ events from Monte Carlo (MC) samples in reconstructing the relevant final states is compared to that of the anti-$k_T$algorithm. Measures of infra-red (IR) safety, both soft and collinear, are also studied. We find that the ability of spectral clustering in reconstructing mass peaks is comparable to that of anti-$k_T$under realistic acceptance and selection criteria. Unlike the standard approach in which the jet cone size needs to be adapted to the kinematics of the events under study, the new algorithm does not require an adjustment of its parameter settings for the processes analysed.

*E-mail: `sd@ecs.soton.ac.uk`
†E-mail: `hadh1g17@soton.ac.uk`
‡E-mail: `b.ford@soton.ac.uk`
§E-mail: `stefano@phys.soton.ac.uk`
¶E-mail: `claire.shepherd@stfc.ac.uk`

1 Introduction

The preferred choice for jet clustering in the context of hadron collider physics tends to be one of three algorithms: the anti-k_T [6], the Cambridge-Aachen [10, 26] or the k_T one [11], that have eventually replaced cone algorithms, which had seen widespread use prior to their advent, all of these having seen their origin in e^+e^- physics, see Refs. [25, 3, 7, 17]. They have been the default choice for some time because they have a number of desirable properties. They are infrared safe, excellent implementations of them are publicly available (see `FASTJET` [4]) and they are flexible enough to capture many different jets signals with minimal parameter changes. These algorithms are recursive (or iterative) and agglomerative. A recursive algorithm is well suited to clustering objects when the number of groups is not known from the outset. Agglomerative algorithms are easier to design in a manner that is infrared safe, as collinear particles tend to combine early due to their small angle while soft particles tend to combine with hard particles rather than combining among themselves.

Jet definition precedes further algorithmic methods to extract useful physical quantities. Finding an alternative clustering method that compares favourably to these popular jet algorithms and which offer additional features for further analysis is a useful goal. Success in obtaining clusters based on informative transformations of the data offers the possibility of exploiting such representations. In this paper, we use Laplacian eigenmaps [2] to represent the particles in an event, a procedure employed in applications such as image segmentation [21] and called spectral clustering [18]. Spectral clustering has had success also in other physics contexts, such as to identify the motion of vortices [12] in fluid dynamics and determining the correct number of clusters. Furthermore, to reduce the risk of blackouts, power grids may be subdivided into 'islands', which are electromechanically stable regions with minimum load shedding. The ideal location of such islands is found by minimising the power flow between them using spectral clustering as shown in [13]. A hierarchical, agglomerative algorithm for the same was introduced in [20] ~~that spectral clustering can produce a good solution in less time than other algorithms commonly used for such a problem (e.g., those using combinatorial optimisation approaches)~~. This agglomerative approach is what we show to be suitable for the context of jet physics in this paper.

~~To our knowledge a spectral clustering algorithm has not yet been applied to the definition of jets, however, given its recursive and agglomerative form, we will show that it is indeed suitable to such a physics context. However, we mention already that ours is a non-standard approach to spectral clustering. In fact, while the embedding step relaxes an optimisation objective the agglomerative step does not (this is described in depth later on).~~

The plan of this paper is as follows. In the next section, we will introduce the fundamentals of the theory of spectral clustering. In the following one, we will describe the details of the specific method that we have applied. The numerical results will then follow. Finally, we will draw our conclusions.

2 Theory of spectral clustering

Collimated emissions of particles are clustered by jet algorithms. A representation of observable particles that preserves and accentuates local information motivates the Laplacian eigenmap [2] and spectral clustering [18]. Spectral clustering is a method by which a set of points are represented in a new space, called the embedding space, in which they can be easily clustered. Coordinates of the points in the embedding space are expressed in terms of the eigenvectors and eigenvalues of an associated Laplacian matrix, hence the name.

The particles in an event are described first as nodes of a graph and edges capturing a notion of similarity between them. The theory behind the construction of the embedding space is a relaxation of criteria that would precisely partition nodes into separate disconnected subgraphs. ~~This criteria can then be relaxed to permit some connections between subgraphs.~~ An excellent description can be found in [15]; a short summary is given here.

At the start we have a group of points with coordinates, which should be split into a number c of predetermined clusters. Applying the spectral clustering method requires making these points into a graph. A simple way to do this would be to consider the points to be the nodes of a fully connected graph. The vertex of the graph joining node (or point) i and j has weight $a_{i,j}$, which should grow with the probability of i and j being in the same group.

The initial aim is to identify which of the components each point belongs to, by sorting the graph into subgraphs, G_k, where $k = 1 \dots c$. Minimising the NCut objective is a function that captures this aim, where

$$\mathrm{NCut} = \frac{1}{2}\sum_k \frac{W(G_k, \bar{G_k})}{\mathrm{vol}(G_k)}, \tag{1}$$

where $W(G_k, \bar{G_k})$ is the sum of all the vertex weights that must be dropped to separate the cluster G_k from the rest of the graph, $\bar{G_k}$. So that $W(G_k, \bar{G_k}) = \sum_{i \in G_k, j \in \bar{G_k}} a_{i,j}$. In the denominator $\mathrm{vol}(G_k) = \sum_{i \in G_k} \sum_j a_{i,j}$, the sum of all affinities connecting to a point in G_k. This denominator is used to penalise forming small clusters.

In order to determine which point will go in which G_k, a set of indicator vectors must be found. Membership of cluster G_k will be recorded in the indicator vector h_k:

$$h_{i,k} = \begin{cases} 1/\sqrt{\text{vol}(G_k)} & \text{if point } i \in G_k, \\ 0 & \text{otherwise,} \end{cases}. \tag{2}$$

To find these indicator vectors the graph is represented by the graph Laplacian, L, a square matrix with as many rows and columns as there are points. To construct this Laplacien we define two other matrices; an off diagonal matrix $A_{i,j} = (1 - \delta_{i,j})a_{i,j}$ and a diagonal matrix $D_{i,j} = \delta_{i,j} \sum_q a_{i,q}$. Then the symmetric Laplacian can be simply written as;

$$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} \tag{3}$$

Notice that this is a real symmetric matrix and, therefore all its eigenvalues are real. Considering just one cluster, G_k, when the Laplacian is multiplied by its indicator vector, the result is the term that NCut seeks to minimise for that cluster.

$$h_k' L h_k = \frac{1}{\text{vol}(G_k)} \sum_{i \in G_k, j \in G_k} \left(\delta_{i,j} \sum_l a_{l,i} - a_{i,j} \right) = \frac{W(G_k, \bar{G_k})}{\text{vol}(G_k)} \tag{4}$$

To obtain the sum of all the terms, stack the indicator vectors into a matrix, $h_k' L h_k = (H' L H)_{kk}$, and the NCut aim described earlier becomes the trace,

$$\text{NCut}(G_1, G_2, \ldots G_n) \equiv \frac{1}{2} \sum_{k=1}^{n} \frac{W(G_k, \bar{G_k})}{\text{vol}(G_k)} = \text{Tr}(H' L H), \tag{5}$$

where $H'H = I$. This is still an NP hard problem, however if we relax the requirements made on h in Eqn. 2, allowing the elements of h to take arbitrary values, then the Rayleigh-Ritz theorem provides a solution. Trace minimisation in this form is done by finding the eigenvectors of L with smallest eigenvalues. Due to the form of the Laplacian, there will be an eigenvector with components all of the same value and its eigenvalue will be 0. This corresponds to the trivial solution of considering all points to be in one group. The next c eigenvectors of L, sorted by smallest eigenvector, are the indicator vectors needed to allocate points to c clusters.

These indicator vectors are then used to determine position of the points in the embedding space. Each indicator vector has as many elements as there are points to be clustered, so the coordinates of a point are the corresponding elements or the indicator vectors. This is all the information the theory of spectral clustering provides. The steps required to make use of this information are not dictated by the theory, and they must be carefully selected to respect the physics.

Using the positions in embedding space the points can be gathered agglomeratively, so that we do not need to chose a predetermined number of clusters.

2.1 Distance in the embedding space

When the relaxed spectral clustering algorithm is used to create an embedding space, points in a group will not be at exactly the same coordinates. Each point can be seen as a vector, the direction of this vector indicates the group to which this point should be assigned. The magnitude indicates the confidence with which the assignment is made. Changes in magnitude cause the Euclidean distance between the corresponding points to grow. An angular distance is appropriate, though. The angular distance will grow when the eigenvectors indicating the point have less overlap and this is what should be measured.

2.2 Information in the eigenvalues

When the clusters in the data are very clear, the situation is closer to the ideal one and the eigenvalues will be closer to 0. The smaller an eigenvalue is, the more like a perfect indicator vector the corresponding eigenvector is. It is possible to make use of this information.

In a traditional application of spectral clustering, the number of clusters desired, c, is predetermined. The embedding space is created by taking c eigenvectors with smallest eigenvalues, excluding the trivial eigenvector. The embedding space then has c dimensions. This follows from a relaxation of the concept of indicator vectors.

When forming jets we do not know from the outset how many clusters to expect in the dataset, so the number of eigenvectors to keep is not clear. While we could chose a fixed, arbitrary number of eigenvectors, this is suboptimal. A better approach is to take all non-trivial eigenvectors corresponding to eigenvalues smaller than some limiting number, λ_{limit}. For a symmetric matrix the eigenvalues will be $0 < \lambda < 2$, so $\lambda_{\text{limit}} = 0.5$ would be a reasonable choice. Then, the number of dimensions in the embedding space will vary, according to the number of non-trivial eigenvectors with corresponding $\lambda < \lambda_{\text{limit}}$.

There is one more manipulation from the information in the eigenvalues. The dimensions of this embedding space are not of equal importance, those with higher eigenvalues being less interesting. This can be accounted for by dividing the eigenvector by some power, β, of the eigenvalue.

Let the eigenvectors for which $\lambda < \lambda_{\text{limit}}$ be

$$L_{i,j}(h_n)_j = \lambda_n(h_n)_i. \tag{6}$$

Then, the coordinates of the j^{th} point in the c dimensional embedding space become $m_j = \left(\lambda_1^{-\beta}(h_1)_j, \ldots \lambda_c^{-\beta}(h_c)_j,\right)$. In effect the n^{th} dimension is compressed by a factor λ_n^{β}, so the larger λ_n the greater the compression.

2.3 Stopping conditions

If a recursive algorithm is to be chosen, like in the generalised k_T algorithm, a stopping condition is needed. A stopping condition based on smallest distance between points in the embedding space does not prove to be stable, as the distribution in the number of dimensions in the embedding space changes sharply from event to event.

The average distance between points is more stable. If this were used in physical space it would force roughly the same number of clusters to form each time, however, the variable number of dimensions in the embedding space is now an advantage. The clearer information found about clusters in the points the more dimensions the embedding space will contain, as described in section 2.2.

Say, the data contains two points that would form a good cluster. If those two points are combined into one, that cluster is complete, fewer clusters remain unfinished and the information for clustering the resulting points will be reduced. When the embedding space is recalculated for the new points, it will likely have fewer dimensions. In a space with fewer dimensions the mean distance between the points naturally falls. Thus, the mean distance in the embedding space is a good indicator of the number of unfinished clusters available. In short, the mean distance in the embedding space makes a natural cut-off.

3 Method

In this section the methodology is covered in four parts. Firstly, the practical procedure chosen in this work for applying spectral clustering is given. Secondly, choices and interpretations for the variable parameters in this algorithm are given. Thirdly, the datasets against which this will be measured are specified. Fourthly, the procedure for checking IR safety is described.

For clarity, note that the variable pseudorapidity is never used in the algorithms

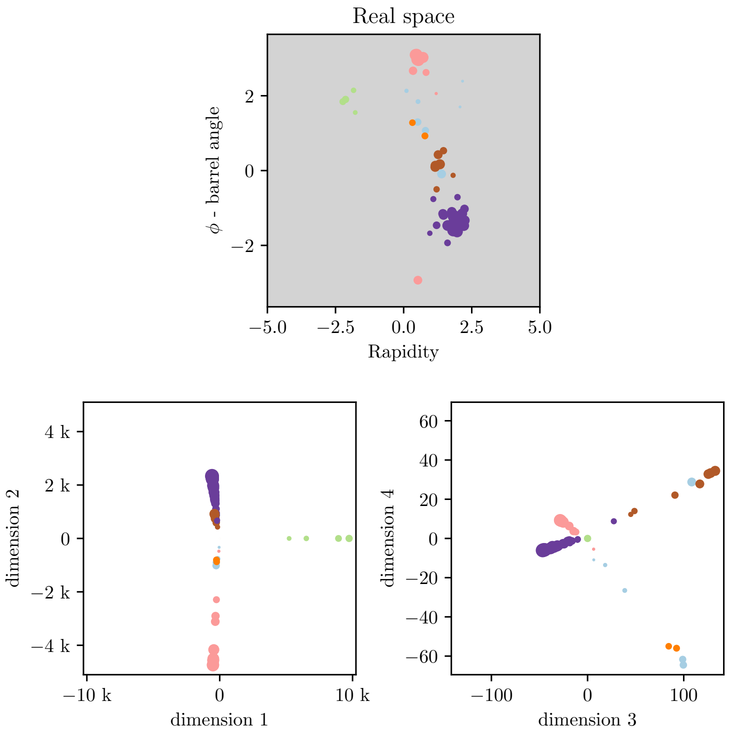

Figure 1: A single event and its embedding space, as created by spectral clustering. At the top the grey plot shows the particles in the event as points on the unrolled detector barrel. The colour of each point indicates the shower it came from. The lower two plots show the first 4 dimensions of the embedding space, and the location of the points within the embedding space.

proposed; all references to rapidity y correspond to

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \tag{7}$$

Besides rapidity, barrel angle, ϕ, is also used as a coordinate. The barrel angle is the angle of the particle in the plane perpendicular to the beam. Barrel angle and rapidity form an orthogonal coordinate system.

3.1 Spectral clustering algorithm

For every simulated event the following process is used to identify the jets [1]. To begin with, relevant cuts are applied to the particles to simulate the detector's reconstruction capability. (These are described in detail in section 3.3.) Then all particles are declared pseudojets and given an index, $j = 1 \dots n$, with no particular order. The algorithm is agglomerative, recursively selecting pairs of pseudojets to merge; hence, the first iteration step is labelled $t = 1$.

When the two pseudojets to be merged, i and j, have been identified they are combined using the E-scheme. The E-scheme forms a new pseudojet by summing the 4-momentum of the two join pseudojets; $p(t+1)_l = p(t)_i + p(t)_j$. The steps used to select two pseudojets to merge proceed as follows;

1. The pseudojets are used to form the nodes of a graph, the edges of which will be weighted by some measure of proximity between the particles called affinity. To obtain an affinity, first a distance is obtained. Between pseudojets i and j this would be $d(t)_{i,j} = \sqrt{(y(t)_i - y(t)_j)^2 + (\phi(t)_i - \phi(t)_j)^2}$ where $y(t)_j$ is the rapidity of pseudojet j at time step t and $\phi(t)_j$ is the barrel angle, likewise for i. No p_T dependence is used, as customary in many traditional jet clustering methods.

2. The affinity must increase as pseudojets become more similar, whereas the distance will shrink. We chose $a(t)_{i,j} = \exp(-d(t)^{\alpha}_{i,j}/\sigma_v)$, where $\alpha = 2$ is the standard Gaussian kernel as used in [2]. Distances much larger than σ_v are only allowed very small affinities, thus less influence over the clustering.

3. Pseudojets that are far apart have low affinity, hence are unlikely to be good candidates for combination. Removing these affinities reduces noise. A fixed number, k_{NN}, of neighbours of each pseudojet is preserved while all other affinities are set to zero. Thus, in a group with more than k_{NN} pseudojets, each pseudojet has at least k_{NN} non-zero affinities with other pseudojets.

4. These affinities allow the construction of the symmetric normalised Laplacian; which is proportional to $-a(t)_{i,j}$ in the i^{th} row and j^{th} column and exactly 1 on the diagonal. For ease of notation, let $z(t)_j$ be a measure of the size a pseudojet j contributes to a cluster, with $z(1)_j = \sum_k a_{i,k}$. Then define square matrices $A(t)_{i,j} = (1 - \delta_{i,j})a(t)_{i,j}$ and $Z(t)_{i,j} = \delta_{i,j} z(t)_i$ The Laplacian can now be written as

$$L(t) = Z(t)^{-\frac{1}{2}}(Z - A)Z(t)^{-\frac{1}{2}} \qquad (8)$$

[1] Code available at https://github.com/HenryDayHall/jetTools

After each step this Laplacian shrinks by one row and column. When two pseudojets have been combined, instead of calculating z_j as the sum of the affinities of the combined pseudojet, the new z_j is the sum of the two previous z_j's. For example, if pseudojets 1 and 2 from $t = 1$ are to be combined to make pseudojet 1 in $t = 2$, then $z(2)_1 = z(1)_1 + z(1)_2$ rather than the sum of affinities between the new pseudojet 1 and other pseudojets in step $t = 2$. This condition is seen to be required for IR safety.

As such, after the first time step, L will no longer be a proper graph Laplacian. Its rows and columns do not sum to zero. However, this new L appears to maintains similar behaviour to a propper Laplacian in the embedding space it creates.

5. The eigenvectors of $L(t)$, (q the eigenvalue index)

$$L(t)h(t)_q = \lambda(t)_q h(t)_q, \; q = 1, \ldots, c \tag{9}$$

are used to create the embedding of the pseudojets. The eigenvector corrisponding to the smallest eigenvalue represents the trivial solution, that places all points in the same cluster (see Sec. 2). All non-trivial eigenvectors, corresponding to eigenvalues less that an eigenvalue limit $\lambda(t)_c < \lambda_{\text{limit}} < \lambda(t)_{c+1}$ are retained. See section 2.2. NOTE: Eigenvalues $0 \leq \lambda(1) \leq 2$. However, $\lambda(t)$ for $t > 1$ are bounded but in a different range.

6. A eigenvector is divided by the corresponding eigenvalue raised to β. This acts to compress the dimensions that hold less information, again, see section 2.2. The embedding space can now be formed. The eigenvectors have as many elements as there are pseudojets and the coordinates of the j^{th} pseudojet at time step t are defined to be $m(t)_j = \left(\lambda_1(t)^{-\beta} h_1(t)_j, \ldots \lambda_c(t)^{-\beta} h_c(t)_j\right)$.

7. A measure of distance between all pseudojets in the embedding space is calculated. In the embedding space angular distances are most appropriate (see section 2.1):

$$d'(t)_{i,j} = \arccos \left| \frac{m(t)_i \cdot m(t)_j}{\|m(t)_i\| \|m(t)_j\|} \right| . \tag{10}$$

where $\|m\|$ is the (Euclidean) length of m.

8. Provided the mean of this distance is less than R, that is,

$$\frac{2}{c(c-1)} \sum_{i \neq j} d'(t)_{i,j} < R, \tag{11}$$

then the two pseudojets that have the smallest embedding distance are combined. NOTE: there are $\frac{c(c-1)}{2}$ possible pairs, where c is the number of pseudojets remaining. (Reasons for this stopping condition are given in section 2.3.)

~~The two pseudojets that have been combined are removed and replaced with one pseudojet. In physical space the new pseudojet is created by summing the respective four-momenta of the removed pseudojets. Once two pseudojets have been combined in physical space, the embedding space is recalculated from step 1, to begin time step $t+1$. There will be one fewer row and column in the Laplacian of step $t+1$.~~

When the mean of the distances in the embedding space rises above R, then all remaining pseudojets are promoted to jets. Jets with less than 2 tracks are removed and their contents considered noise. Further cuts may then be applied as described in section 3.3.

These steps will form a variable number of jets from a variable number of particles. An example of the constructed first embedding space is shown in Fig. 1. This illustrates how the embedding space highlights the clusters.

3.2 Tunable parameters

Unlike most Machine Learning (ML) techniques, spectral clustering does not have large arrays of learnt parameters. The parameters for the clustering are a small, interpretable set. Appropriate values were chosen by performing scans and observing the influence of changes to the parameters on jets formed.

In section 3.1, 6 parameters are named: σ_v, α, k_{NN}, λ_{limit}, β and R. These parameters have a range of values for which sensible results are obtained. The interpretation of these parameters is as follows.

- σ_v: introduced in step 2, this is a scale parameter in physical space. The value indicates an approximate average distance for particles in the same shower, or alternatively, the size of the neighbourhood of each particle. It is closely tied to the stopping parameter for the generalised k_T algorithm, R_{k_T}, they both relate to the width of the jets formed. It should take values on the same order of magnitude as R_{k_T}, ~~, i.e., of $\mathcal{O}(0.1)$~~.

- α: also introduced in step 2, this changes the shape of the distribution used to describe the neighbourhood of a particle. Higher values reduces the probability of joining particles outside σ_v; $\alpha = 2$ defines a Gaussian kernel.

- k_{NN}: introduced in step 3, this dictates the minimum number of non-zero affinities around each point. Lower values create a sparser affinity matrix, reducing noise at the potential cost of lost signal. Values above 7 are seen to have little impact.

- λ_{limit}: introduced in step 5, is a means of limiting the number of eigenvectors used to create dimensions in the embedding space. Only eigenvectors corresponding to eigenvalues less than λ_{limit} are used. Thus, the number of dimensions in the embedding space can be increased with a larger λ_{limit}. However, as the eigenvalues will be influenced by the number of clear clusters available, there will not be the same number of dimensions in each event. (This is discussed in section 2.2.) For a symmetric Laplacian the eigenvalues are $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \lambda_n \leq 2$, and λ_k is related to the quality of forming k clusters [**LeeGharanTrevisan2014**], so values of $\lambda_{limit} < 1$ are sensible choices.

- β: introduced in step 6, it accounts for variable quality of information in the eigenvectors, as given by their eigenvalues, in such a way that the dimensions of the embedding spaces corresponding to higher eigenvalues ~~with lower quality information~~ are compressed. ~~The larger the value of β the more dimensions with lower quality information are compressed.~~ in such a way that the dimensions of the embedding spaces with lower quality information are compressed. The larger the value of β the more dimensions with lower quality information are compressed. (This is discussed in section 2.2.)

- R: is introduced in step 8, it determines the expected spacing between jets in the embedding space. As the number of dimensions in the embedding space grows with increasing number of clear clusters, it will not result in the same or similar number of clusters each time.

To investigate the behaviour of the clustering when the parameters change, scans where performed. On a small sample of 2000 events the clustering is performed with many different parameter choices. With the aid of MC truth information a metric of success can be created. For each object we wish to find (e.g., a b-quark) the MC truth can reveal which of the particles that are visible to the detector have been created by that object. In many cases, a particle seen in the detector will have been created by two objects, such as a particle coming from an interaction between a $b\bar{b}$ pair, in these cases both objects are considered together. The complete set of visible particles that came from these objects could be referred to as their descendants. The aim in jet clustering is to capture only all of the descendants in the same number of jets as there were objects that created them. So the descendants of a $b\bar{b}$ pair should be captured in exactly 2 jets.

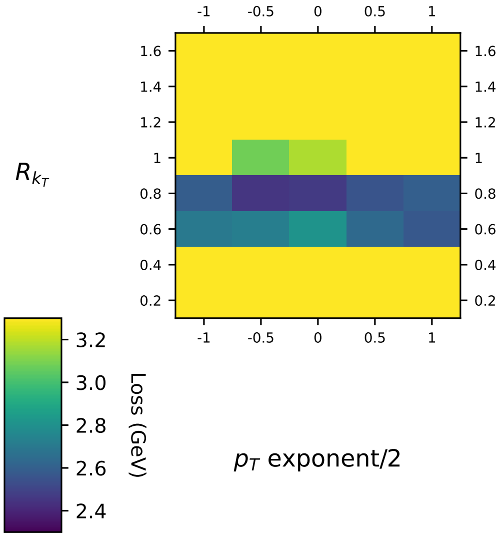

Figure 2: The generalised k_T algorithm has 2 parameters that can be varied. The stopping condition, R_{k_T}, and a multiple for the exponent of the p_T factor. When the exponent of the p_T factor is -1 the algorithm becomes the anti-k_T algorithm. Here, the "Loss", as described in Eq. (12), is shown as a colour gauge for a number of parameter combinations. Please change R_{kt} to R_{k_T} in the y label and add the label Loss near the colour gauge H. done

There are two ways a jet finding algorithm can make mistakes in this task: the first is to omit some of the descendants of the objects being reconstructed, causing the jet to have less mass than it should; the second is to include particles that are not in the descendants of the objects being reconstructed, such as initial state radiation or particles from other objects, causing the jet to have more mass than it should. The effects of these mistakes will cancel in the jet mass, but they are both still individually undesirable, so separate metrics are made for each of them. The first is "signal mass lost", the difference between the mass of the jets and the mass they would have had if all they contained all descendants of the object being reconstructed. The second is "background contamination", the difference between the mass of jets and the mass they would have if they did not contain anything but descendants of the objects being reconstructed. A loss function is constructed as a weighted euclidean combination of these two;

$$\text{Loss} = \sqrt{w\,(\text{Background contamination})^2 + (\text{Signal mass lost})^2}. \tag{12}$$

Where w is a weighting used to alter the preference for suppressing signal mass lost verses reducing background contamination. When applying an anti-k_T algorithm, increasing R_{k_T} will result in lower signal mass loss, in exchange for a higher background contamination. The standard choice for this process is $R_{k_T} = 0.8$. This value of R_{k_T} slightly prefers suppressing signal mass lost over background contamination, to create the clearest mass peaks. To make the loss reflect this we chose $w = 0.73$. Please explain the origin of that 0.73 H. improved?

An example of this scan for the generalised k_T algorithm is given in Fig. 2. It can be seen that, while good results are possible with many values of the p_T exponent, R_{k_T} must fall in a narrow range. We thus deem this choice of stopping condition, $R_{k_T} = 0.8$, to be rather fine-tuned.

For spectral clustering there are more than 2 variables to deal with, so a set of two dimensional slices are extracted. These slices have been chosen to include the best performing combination. As can be seen in Fig. 3, the parameters choices are not fine-tuned. That is, unlike the anti-k_T algorithm, there is flexibility in all parameter choices. For example, it can be seen that some parameters, such as α, k_{NN}, β and λ_{limit} are relatively unconstrained, yielding good results for a wide range of values. Even when R and, especially, σ_v yield some large signal losses, say, for $R = 1.22$ or 1.3 and $\sigma_v = 0.05$, this happens is very narrow ranges. For definiteness, the parameters used in the remainder of this work are $\alpha = 2.$, $k_{\mathrm{NN}} = 5$, $R = 1.26$, $\beta = 1.4$, $\sigma_v = 0.15$ and $\lambda_{\mathrm{limit}} = 0.4$.

3.3 Particle data

To evaluate the behaviour of the spectral clustering method four datasets are used[2], all produced for the Large Hadron Collider (LHC).

1. Light Higgs A SM-like Higgs boson with a mass 125 GeV decays into two light Higgs states with mass 40 GeV, which in turn decay to $b\bar{b}$ quark pairs. That is, the process is $pp \to H_{125\,\mathrm{GeV}} \to h_{40\,\mathrm{GeV}} h_{40\,\mathrm{GeV}} \to b\bar{b}b\bar{b}$, simulated at Leading Order (LO).

2. Heavy Higgs A heavy Higgs boson with a mass 500 GeV decays into two SM-like Higgs states with mass 125 GeV, which in turn decay to $b\bar{b}$ quark pairs. That is, the process is $pp \to H_{500\,\mathrm{GeV}} \to h_{125\,\mathrm{GeV}} h_{125\,\mathrm{GeV}} \to b\bar{b}b\bar{b}$, simulated at LO.

3. Top A $t\bar{t}$ pair decays semileptonically, i.e., where one $W^\pm$ decays to a pair of quark jets jj and the other into a lepton-neutrino pair $\ell\nu_\ell$ ($\ell = e, \mu$). That is, the process is $pp \to t\bar{t} \to b\bar{b}W^+W^- \to b\bar{b}jj\ell\nu_\ell$, simulated at LO. (Note that, here, $m_t = 172.6$ GeV and $m_W = 80.4$ GeV.)

4. 3-jets For the purpose of checking IR safety, we have used three-jet events, this being a rather simple configuration where IR singularities could be observed. That is, the process is $pp \to jjj$, simulated at both LO and Next-to-LO (NLO).

Using MadGraph [1] to generate the partonic process and Pythia [24] to shower, $\mathcal{O}(10^5)$ of each of these processes are generated. A full detector simulation is not used; instead, cuts on the particles are imposed to approximate detector resolution, as detailed below. Explain how the detector was simulated.H. does this work?

The Center-of-Mass (CM) energy used is $\sqrt{s} = 13$ TeV.

Each event also contains (hard) Initial State Radiation (ISR) and soft QCD dynamics from beam remnants, i.e., the Soft Underlying Event (SUE). There is no pileup or multiparton interactions in the datasets. What about MPIs and pile-up?H. we have neither

Each of these datasets requires different cuts, both at the particle level, to simulate detector coverage, and at the jet level, to select the best reconstructed events. The cuts on each dataset are as follows.

[2] The first two uses a 2-Higgs Doublet Model (2HDM) setup as described in Ref. [8] while the last two are purely Standard Model (SM) processes. Notice that all unstable objects are rather narrow, including the Beyond the SM (BSM) Higgs states [16, 9], so that we have neglected interference effects with their irreducible backgrounds.

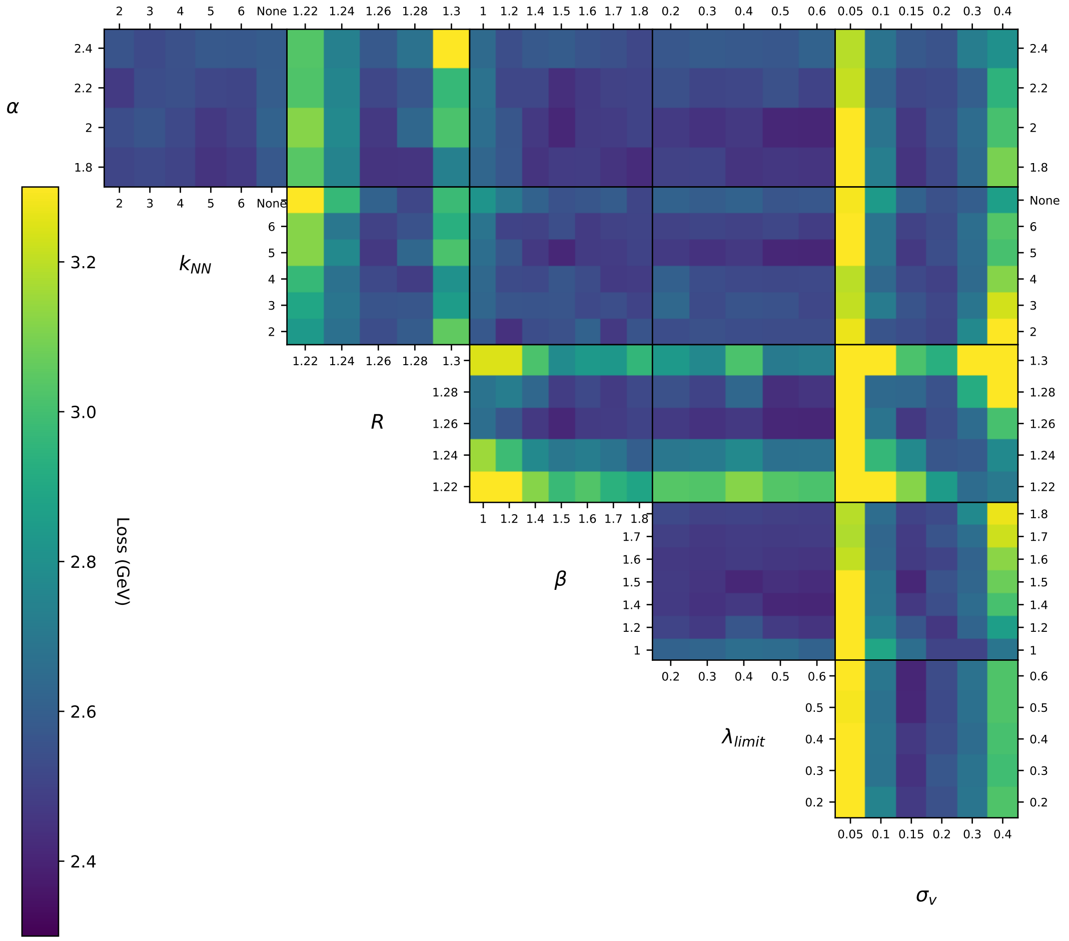

Figure 3: There are 6 free parameters in spectral clustering. Here, the "Loss", as described in Eq. (12), is shown for reasonable parameter ranges chosen either by convention (e.g., α is typically 1 or 2) or according to physical scales (e.g., σ_v is of order 0.1). Please remove the equation next to the colour gauge and just leave Loss in horizontal H. will do

1. The reconstructed particles are required to have rapidity $|\eta| < 2.5$ and (transverse momentum $p_T > 0.5$ GeV. These cuts are likely to remove the majority of the radiation from beam remnants and reduce the radiation from ISR. The b-jets are required to have $p_T > 15$ GeV, which is possibly lower than is realistic [8], but

it leaves a larger number of events to compare the behaviour of jet clustering algorithms.

2. The reconstructed particles are required to have $|\eta| < 2.5$ and $p_T > 0.5$ GeV. The b-jets are required to have $p_T > 30$ GeV, which is realistic for efficient b-tagging performance and further reduces ISR and the SUE.

3. The reconstructed particles are required to have $|\eta| < 2.5$ and $p_T > 0.5$ GeV. The event is required to have $p_T^{\text{miss}} > 50$ GeV, where $p_{T,\text{miss}}$ is the missing transverse momentum due to the neutrino. The lepton in the event must have $|\eta| < 2.4$. If the lepton is a muon then its p_T must be > 55 GeV. If the lepton is an electron and it is isolated (as defined in [23]) then its p_T must be > 55 GeV, if it is not isolated then $p_T > 120$ GeV. The reconstructed jets must have $p_T > 30$ and $|\eta| < 2.4$. Finally, the lepton must be separated from the closest jet by at least $\sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.4$ or $p_T^{\text{relative}} > 40$ GeV. These cuts are copied from [22].

4. The only restriction on the particles is that the rapidity must be < 2.5. There are no cuts on the jets. While unrealistic, since issues of IR safety are emphasised at low p_T, to highlight this we abandon all p_T cuts.

The Higgs boson cascade datasets have the desirable property of creating b-jets with different kinematics: while in case 1 we may expect some slim jets (as on average they are rather stationary, because of the small mass difference between $H_{125\,\text{GeV}}$ and $H_{40\,\text{GeV}}$) in case 2 we may see mainly fat jets (owing to the boost provided by the large mass difference between $H_{500\,\text{GeV}}$ and $H_{125\,\text{GeV}}$). Mass reconstruction requirements for the Light Higgs and Heavy Higgs follow the same logic. In order to reconstruct a Higgs decaying directly to a pair of b-quarks, we require a separate jet tagged by each b-quark, that is, two jets are required, each tagged by a b-quark from that Higgs. To reconstruct a Higgs that decays to a pair of child Higgs particles, we require both child Higgs bosons have been reconstructed, that is, all four b-jets are found. In the case of the Top events three masses can be reconstructed from jets, the hadronic W, the hadronic top and the leptonic top. The hadronic W is reconstructed if both of the quarks it decayed to have tagged jets; they are permitted to tag the same jet, so the hadronic W can be reconstructed from one or two jets. The hadronic top is reconstructed if the hadronic top is reconstructed and the b-quark from the hadronic top has tagged a jet, so the correct b-jet is required in addition to the requirements on the W. The leptonic top is reconstructed if the b-quark from the top decay tags a jet, and the missing momentum calculation which reconstructs the leptonic W yields a real mass. If the missing mass calculation for the mass of the leptonic W yields two real masses, the one closes to the

W mass is selected. Is this correct?H. it's a little different. I have also moved it after the particle and jet cuts, as it occurs afterwards.

We now proceed to compare spectral to anti-k_T clustering and we start from testing IR safety of the former, while this is a well-known feature of the latter. We will then move on to study Higgs boson and top quark events.

3.4 Determining IR safety

It would be possible to demonstrate IR safety analytically, however, as the environment required for clustering on MC data is already set up, it is more efficient for this study to prove IR safety with such data. This can be done by showing that an IR sensitive variable, for example, the jet mass spectrum, is stable between a LO dataset with no IR singularities and a NLO one which will instead contain IR singularities.

Showing the jet mass spectrum at LO and NLO for a particular configuration, that is, a particular selection of clustering parameters, would allow a comparison that would highlight any differences caused by IR sensitivity. This will be done for illustrative purposes, however, even an IR unsafe algorithm, such as the iterative cone one [5], has some configurations for which these singularities are avoided.

To provide a more global view, a scan of parameter configurations must be compared. Thus, for an unsafe algorithm (such as the iterative cone) the unsafe configuration will be found. It would be cumbersome to compare all these jet mass spectrum by eye, however. Instead, we introduce a summary statistic representing the divergence between two distributions, the Jensen-Shannon score [14].

The Jensen-Shannon score is a value computed between two distributions that increases in magnitude the more these distributions differ. It is a symmetrised variant of the Kullback-Leibler divergence. [Ref needed]H., the definition of JS and KL are both given in [14], along with an explanation of that they are. I think JS is originally from [14], though something similar exists in https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1946.0056. Should I be citing [14] here? The Kullback-Leibler divergence between probability densities p and q can be written as

$$D_{\text{KL}}(p|q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \tag{13}$$

from which the Jensen-Shannon divergence can be written as

$$D_{\text{JS}}(p,q) = \frac{1}{2} D\left(p|\frac{1}{2}(p+q)\right) + \frac{1}{2} D\left(q|\frac{1}{2}(p+q)\right). \tag{14}$$

Here, D_{JS} treats p and q symmetrically and will grow as they become more different. The spectrum of Jensen-Shannon scores will be plotted for a known IR safe clustering algorithm, anti-k_T, a known unsafe clustering algorithm, iterative cone, and the spectral algorithm. If the Jensen-Shannon scores for spectral are consistently small, then it is IR safe.

4 Results

Before the behaviour of the algorithms is analysed, some plots of kinematic variables are shown in Fig. 4. It can be seen that the algorithms do not greatly differ on the kinematics of the events. In particular, spectral clustering does not appear to sculpt any distributions in any of the datasets involving Higgs bosons and top (anti)quarks.

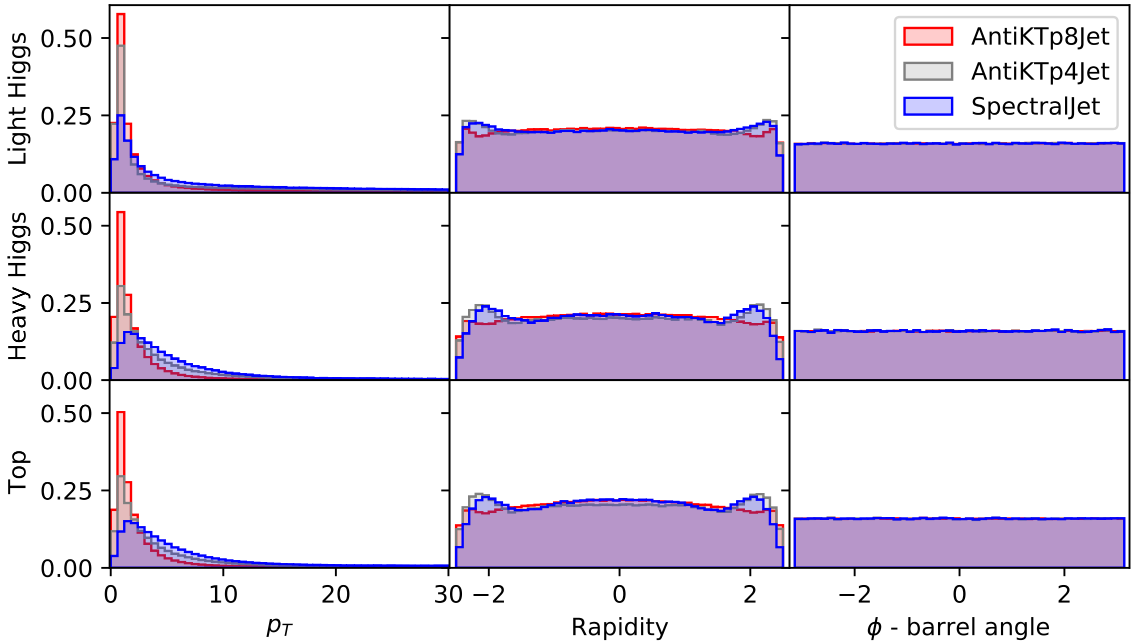

Figure 4: Basic jet variables for each of the analysis datasets and three clustering algorithms. In the first column there is some noticeable differences in the transverse momentum. In the second column the rapidity shows that the algorithms cluster jets at the edge of the barrel slightly differently. In the third column the barrel angle show no noticeable changes.

4.1 IR safety

Shape variables (see the QCD section of Ref. [**Altarelli:116932**] for a useful review), such as jet mass, thrust, sphericity and oblataness, are sensitive to IR divergences. For each configuration of the clustering algorithm we expect an IR safe algorithm to present a stable transition in a shape variable from the LO to NLO datasets, as significant changes in the spectra would indicate sensitivity to soft and collinear radiation. The clustering and evaluation here is done using the 3-jets dataset, as described in Sec. 3.3. Shape variables are calculated from the total momentum of the 4 jets with highest p_T in each event. This comparison is made in Fig. 5. It can be seen in this figure that little difference exists between generalised k_T and spectral clustering, so as to reinforce that they are both IR safe. What data set has been used here? Also, it would be woth to discuss why the distributions are so different for the case of the Mass variable. Finally, is the latter just the invariant mass formed by all tracks/particles in each jet?H. I have now specifed the dataset and the choice of momentum vectors in each event. I also changed the parameters of genkt/spectral used so that they minimc each other. It just wanted a diferent parameter choice.

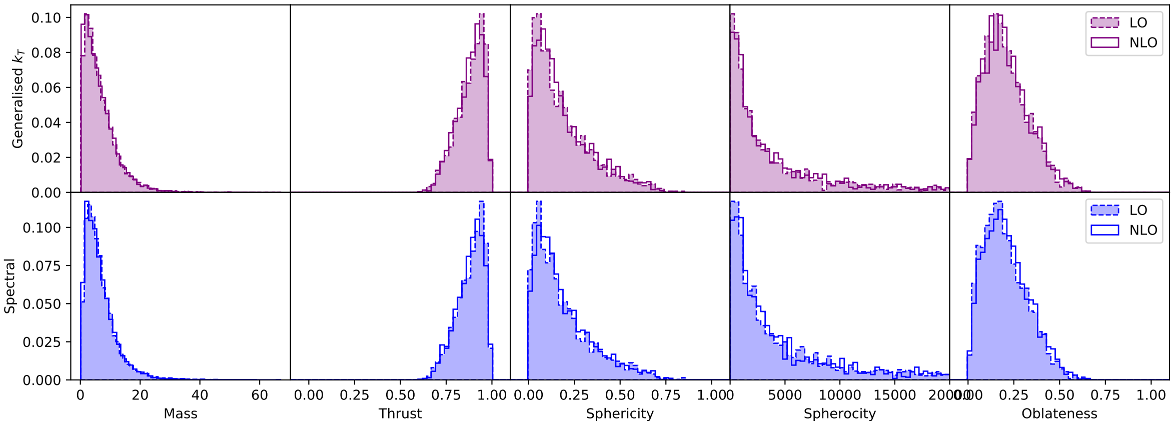

Figure 5: Spectra for jet properties created with LO and NLO datasets. The 4 jets with highest p_T from each event are used in aggregate as an average to form these plots. The columns from left to right are: the jet mass, thrust, sphericity and oblateness. Algorithms where configured (i.e., settings of R) to give sensible results on this dataset, therefore distributions may not represent worst case scenarios.

However, this method of establishing IR safety only looks at one hyperparameter configuration and could be accused of cherry-picking. As described in section 3.4, this can be systematically compared for many hyperparameter configurations by calculating a Jensen-Shannon score for each LO and NLO pair of jet mass spectra. If the Jensen-Shannon metric is low, then the two distributions are similar and appear IR safe. To further clarify the result we include an algorithm known to be IR unsafe, the iterative cone algorithm. The spectral method produces Jensen-Shannon scores very similar to generalised k_T methods. Only the iterative cone one produces high Jensen-Shannon scores thus indicating significant changes between the LO and NLO spectra. This can be seen in Fig. 6. Is this plot done only using the Mass variable of the previous figure? If so, why this choice instead of, e.g., thrust, sphericity, oblataness or others?H. fixed

From the last two figures it is clear that spectral clustering is IR safe, at least, as much as generalised k_T algorithms are. This contrasts with the iterative cone algorithm, for which the jet mass spectra at LO and NLO differ significantly for many configurations. This is not unexpected, as the inputs to the spectral clustering algorithm are the same as for the Cambridge-Aachen one, which is itself IR safe, and the iterative cone has been proved to produce kinematic configurations which are IR unsafe [19] Please refer to some paper illustrating this (by Salam, Seymour, etc.) H. done . However, it is crucial to have such a verification in data, as we have done.

4.2 Mass peak reconstruction

In this section, the anti-k_T algorithm setups with jet radius $R_{k_T} = 0.4$ and $R_{k_T} = 0.8$ are compared to the spectral algorithm specified in section 3.2. The jets are tagged using MC truth. Each of the b-quarks created by a signal particle (either a Higgs boson or a top (anti)quark) tag the closest jet (by using the distance metric $\sqrt{(y_{\text{quark tag}} - y_{\text{jet}})^2 + (\phi_{\text{quark tag}} - \phi_{\text{jet}})^2}$ Do you really mean rapidity here or pseudorapidity? This needs to be clarified throughout as the two are not the same for massive objects H. it is always rapidity, pesudorapidity is never used in this work, is Eqn 7 alright?) provided that the separation between the jet and the quark is no greater than 0.8 according to the distance metric The sentence is unclear improved?. In the case of a $W^\pm$ decay, whatever quarks decay from the $W^\pm$, which may be light quarks or b-quarks, are used to tag jets in the same way. From this point on, only jets tagged this way are considered What about light jets from $W^\pm$ decays? Good point. Fixed..

Firstly, jet multiplicities, that is number of reconstructed jets found per event, are given for both anti-k_T and spectral clustering algorithms. These can be seen for the first three datasets described in section 3.3 in Fig. 7. Herein, it is seen that spectral clustering produces the best multiplicity (i.e., most events where 4 jets are found) for light Higgs events while for the heavy Higgs and top MC samples it creates a multiplicity closer to that of anti-k_T with $R_{k_T} = 0.4$ than $R_{k_T} = 0.8$, the first of these being the

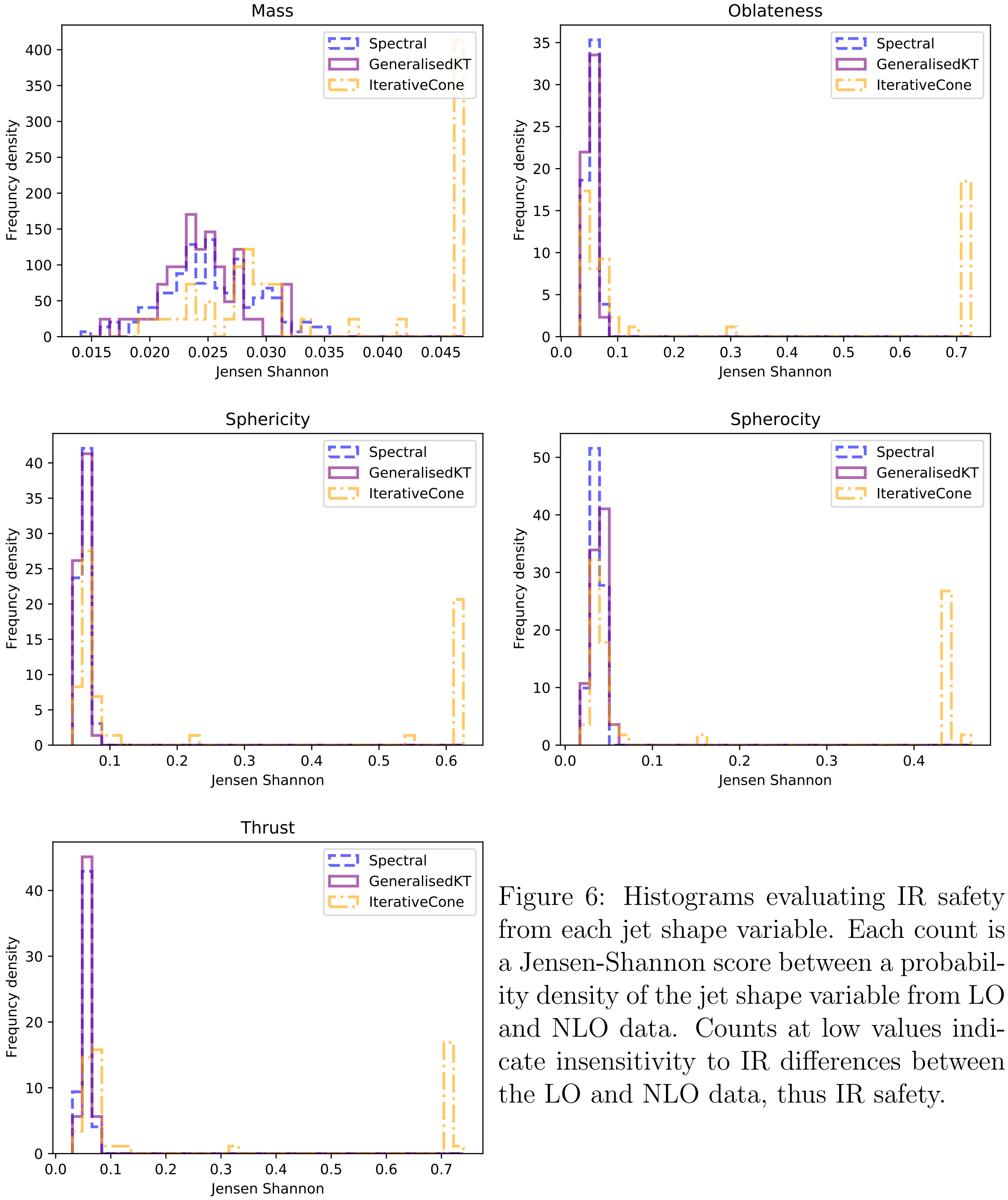

Figure 6: Histograms evaluating IR safety from each jet shape variable. Each count is a Jensen-Shannon score between a probability density of the jet shape variable from LO and NLO data. Counts at low values indicate insensitivity to IR differences between the LO and NLO data, thus IR safety.

best performer of the two. As a result of this study, we remark upon the adaptability of spectral clustering to the different final states without requiring adjusting its parameters, unlike the anti-k_T one. The latter may seem to indicate that 0.4 is the best choice for all datasets, but this is in tension with the fact that different masses from different datasets do require the anti-k_T algorithm to be adjusted, as we shall now see.

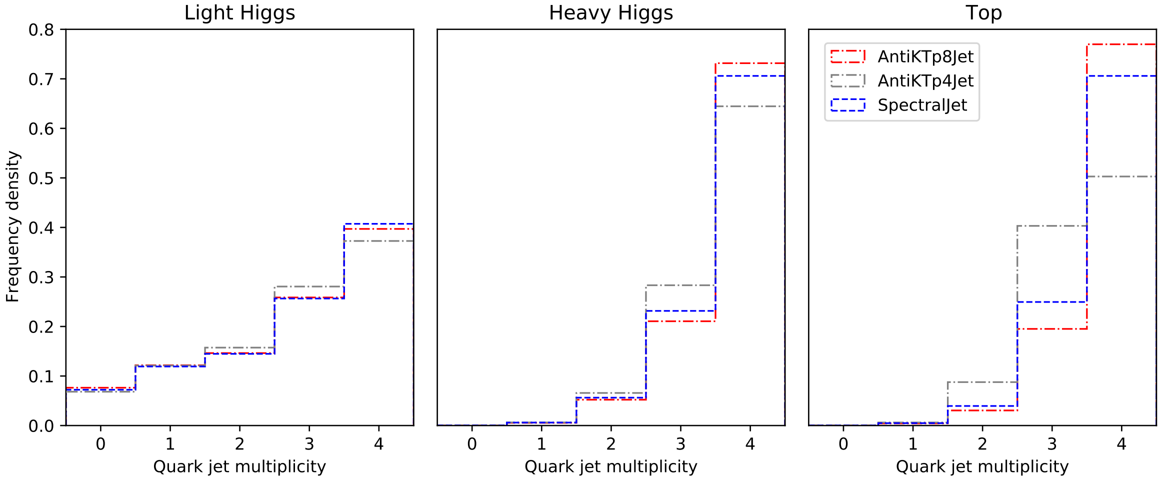

Figure 7: Jet multiplicities for the anti-k_T (for two jet radius choices) and spectral clustering algorithms on the light Higgs, heavy Higgs and top MC samples. For all such datasets, the hard scattering produces 4 partons in the final state, so maximising a multiplicity of 4 jets indicates good performance.

Mass peaks are constructed from the reconstructed jets as well as, for the top sample only, from the lepton and neutrino. Again, the anti-k_T results with $R_{k_T} = 0.4$ and 0.8 are given for comparison. In Fig. 8 three selections are plotted. Firstly, we show events where all 4 b-jets are plotted as total invariant mass of the event, thus reconstructing the mass of the SM Higgs boson. Each event also contains two light Higgs states, though. These are differentiated by the mass of the particles (generating them) that pass the particle cuts, as follows. The light Higgs boson reconstructed from the 2 bi-jet system with more mass visible to the detector is called the “Light Higgs with stronger signal” while the one reconstructed with less mass visible in the detector is called the “Light Higgs with weaker signal” I do not really like this wording, I would suggest instead “Most massive light Higgs” and “Lest massive light Higgs” H., I don’t think the mass of the higgs itself (the displacement offshell) has much influence on which higgs is which.

It’s a reflection of which higgs produces more decay products that pass the particle cuts, whcih reflect the detectors capacity to reconstruct particles according to pt and rapidity. What do you think of; “Higgs producing stronger signal”/“Higgs producing weaker signal”?. The correct jets for each Higgs mass reconstruction are identified using MC truth, so the correct pairings are always made. (If two such di-jet systems are not found the event is not included in the plots). Altogether, it can be seen that spectral clustering forms the sharpest peaks and such peaks are all very close to the correct mass. In fact, its performance is comparable to that of anti-k_T with jet radius 0.8 and is clearly better than the 0.4 option.

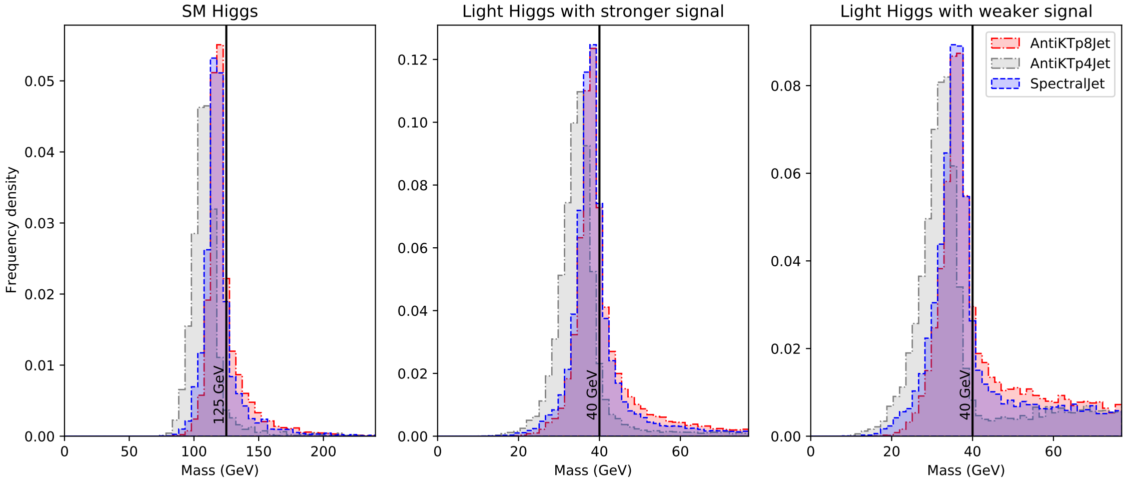

Figure 8: Three mass selections are plotted for the light Higgs dataset. From left to right we show: the invariant mass of the 4 b-jet system, of the 2 b-jet system with heaviest invariant mass and of the 2 b-jet system with lightest invariant mass (as defined in the text). Three jet clustering combinations are plotted as detailed in the legend. The spectral clustering algorithm is consistently the best performer in terms of the narrowest peaks being reconstructed and comparable to anti-k_T with $R_{k_T} = 0.8$ in terms of its shift from the true Higgs mass values, with anti-k_T with $R_{k_T} = 0.4$ always being the outlier. Higgs should be capitalised as Higgs in the top titles, though, see my remark in the body for the actual names H. done

In Fig. 9 the exercise is repeated for the heavy Higgs dataset. All the parameters of spectral clustering are the same as in the light Higgs MC sample yet we note that its performance is still excellent, with very sharp peaks at the correct masses, although the three clustering algorithms are overall much closer in performance. However, recall that, in Fig. 7, it was seen that spectral clustering achieved better multiplicity than anti-k_T with $R_{k_T} = 0.8$ on this dataset. Furthermore, while the multiplicity of anti-k_T with $R_{k_T} = 0.4$ is a little better, the location of all Higgs mass peaks for anti-k_T with $R_{k_T} = 0.4$ is slightly worse. So, we are again driven to conclude that spectral clustering is probably the best performer overall with the added benefit of not requiring any adjustment of its parameters to achieve it.

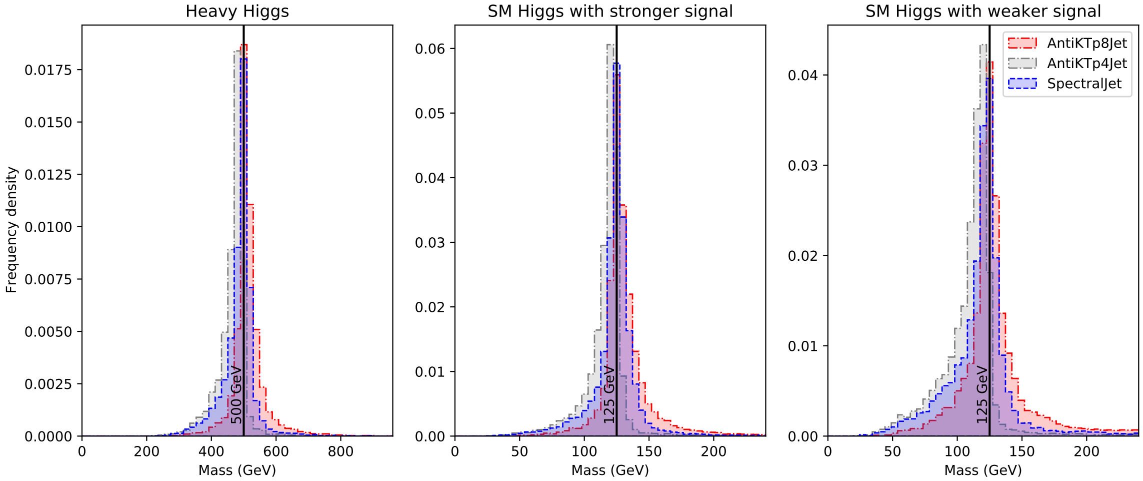

Figure 9: Same as Fig. 8 for the heavy Higgs dataset. Here, the performance of the spectral clustering and anti-k_T (with both 0.4 and 0.8 as jet radiuses) clustering is much closer to each other.

Finally, in Fig. 10, the $W^{\pm}$ and t mass peaks for semileptonic $t\bar{t}$ decays are shown. Three mass reconstructions are given. Firstly, the hadronic W is reconstructed from the jets that come from the quarks it decayed to. Correct decisions about which quarks correspond to which particle in the hard process are made by using information in the Monte Carlo, this is to prevent any mismatching from causing additional complication in evaluating the performance of the clustering. To tag a jet with a quark a distance measure $\sqrt{(y_{\text{quark tag}} - y_{\text{jet}})^2 + (\phi_{\text{quark tag}} - \phi_{\text{jet}})^2}$ is used, and if the distance from the quark to the closest jet is less than 0.8 that jet is tagged by that quark. The W will always decay to a pair of quarks, but both these quarks may be captured in one jet or separate jets. This sentence did not make sense at all H. is this better?. If either of the these quarks are too far away from the closest jet to tag it, that is $\sqrt{(y_{\text{quark tag}} - y_{\text{jet}})^2 + (\phi_{\text{quark tag}} - \phi_{\text{jet}})^2} > 0.8$, then it is not associated with any jet and the hadronic W is not reconstructed. This sentence did not make sense either H. is this better?. The mass of the hadronic top is then reconstructed in events where the hadronic W could be reconstructed and the b-jet from the hadronic top is also found. By

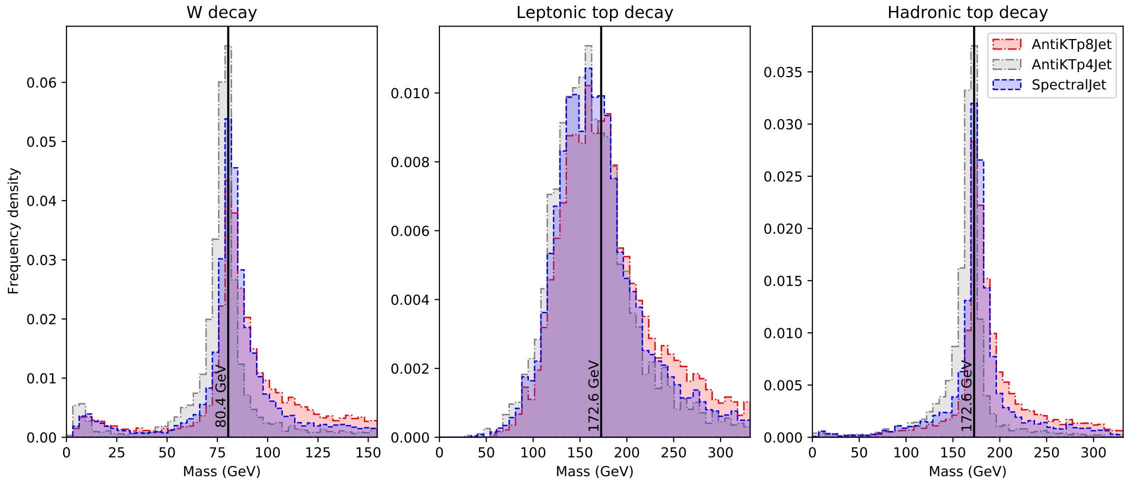

Figure 10: Three mass selections are plotted for the top dataset. From left to right we show: the invariant mass of the light jet system, of the reconstructed leptonic $W^{\pm}$ (as described in the text) combined with a b-jet and of the hadronic $W^{\pm}$ combined with the other b-jet. Three jet clustering combinations are plotted as detailed in the legend. The spectral clustering algorithm consistently outperforms the anti-k_T one with jet radius 0.8 and is slightly worse than the anti-k_T one with jet radius 0.4, but only in tems of sharpness, not location.

MC truth or you pick the b that gives the best m_t? Please clarify H., always MC truth, I will clarify this at the start of the passage.. The leptonic top is then reconstructed in events where b-jet from the top is combined with the reconstructed $W^\pm$ which as decayed leptonically By MC truth or pick the b that gives the best m_t? Please clarifyH. ditto. The leptonic reconstruction of the $W^\pm$ uses the momentum of the electron p_ℓ, the missing transverse momentum p_T^{miss} (identified with that of the neutrino) and the longitudinal neutrino momentum (p_L^ν, which is unknown) in a quadratic equation, $(p_\ell + p_T^{\text{miss}} + p_L^\nu)^2 = m_W^2$, of which only the real solutions are plotted. In this case, it can be seen that spectral clustering is adapting to jets of a different radius. In fact, while before its behaviour had mostly resembled anti-k_T with $R_{k_T} = 0.8$, it has now moved closer to the case with $R_{k_T} = 0.4$. (Semileptonic top events would typically be processed using anti-k_T with $R_{k_T} = 0.4$.) The peaks of spectral clustering are not quite as narrow as those from anti-k_T with $R_{k_T} = 0.4$, but they improve on $R_{k_T} = 0.8$ and their location is substantially correct.

5 Conclusions

Spectral clustering is a popular unsupervised ML algorithm which often outperforms other approaches in many physics contexts, wherein complex multidimensional datasets are reduced into clusters of similar data in fewer dimensions. In performing such a dimensionality reduction, it makes use of the spectrum (eigenvalues) of the similarity matrix of the data. As such, spectral clustering is very simple to implement and can be solved efficiently by standard linear algebra methods. Hence, it is a transparent algorithm with no black box element and all intermediate steps are interpretable. Owing to these features, we have found it to also be a promising new method to apply to jet formation in high energy particle physics events.

For a start, it satisfies the need for IR safety and creates jets with the expected kinematics, as dictated by QCD dynamics. Furthermore, while it has many hyperparameters, they do not appear to be as finely tuned as those of more standard tools, such us sequential (or iterative) generalised k_T algorithms. This can be seen in both parameter scan stability and its adaptability to various datasets, each capturing physics signals embedding heavy objects decaying into lighter ones in very different patterns, all yielding complicated hadronic signatures at the LHC.

The adaptability between datasets is remarkable as a spectral clustering parameter choice tuned on a light Higgs boson cascade gave excellent performance on both a heavy Higgs boson cascade and that of top-antitop pairs decaying semileptonically. In the case of the light Higgs dataset, spectral clustering gave the correct mass peak positions, the narrowest resonant distributions and a jet multiplicity mapping well the partonic one. This would not be surprising as it was tuned for that dataset in the first place. In the case of the heavy Higgs dataset only anti-k_T with $R_{k_T} = 0.8$ and the spectral algorithm gave correct mass peaks but spectral clustering offers considerably better multiplicity rates. This demonstrates that its performance is not dependent on fine tuning its parameters and hence that the algorithm is adaptable to the same final state with different masses involved. Finally, spectral clustering was applied to a dataset with a different final state and for which the ideal jet radius differed, semileptonic decays of top-antitop pairs. Its equivalent parameter σ_v was not allowed to vary to account for this, instead it was applied again with no parameter changes. The algorithm again proved to be adaptable and modified its behaviour to follow that of anti-k_T with $R_{k_T} = 0.4$, the standard choice for this kind of analyses.

In short, spectral clustering is a novel and promising approach to jet formation, which initial development already demonstrates flexibility and excellent performance for numerical analyses at the forefront of collider physics.

6 Acknowledgements

We thank A. Chakraborty, G. Cerro, J. Chaplais, S. Jain and E. Olaiya fur insightful discussions. HAD-H thanks G.P. Salam for useful advice. HAD-H, BF, SM and CHS-T are supported in in part through the NExT Institute. SM is also supported by the STFC Consolidated Grant No. ST/L000296/1. BF is funded by the DISCnet & SEPnet scholarship schemes. We finally acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services, at the University of Southampton, in the completion of this work. Others

References

[1] J. Alwall et al. “MadGraph 5: going beyond”. In: *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128. URL: http://dx.doi.org/10.1007/JHEP06(2011)128.

[2] M. Belkin and P. Niyogi. “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”. In: *Neural Comput.* 15.6 (June 2003), pp. 1373–1396. ISSN: 0899-7667. DOI: 10.1162/089976603321780317. URL: https://doi.org/10.1162/089976603321780317.

[3] S. Bethke et al. “New jet cluster algorithms: Next-to-leading order QCD and hadronization corrections”. In: *Nucl. Phys.* B370 (1992). [Erratum: Nucl. Phys.B523,681(1998)], pp. 310–334. DOI: 10.1016/S0550-3213(98)00219-3,10.1016/0550-3213(92)90289-N.

[4] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet User Manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[5] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-k_t jet clustering algorithm”. In: *Journal of High Energy Physics* 4 (Apr. 16, 2008), p. 13. ISSN: 1126-6708. DOI: 10.1088/1126-6708/2008/04/063. URL: http://stacks.iop.org/1126-6708/2008/i=04/a=063 (visited on 05/02/2018).

[6] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[7] S. Catani et al. “New clustering algorithm for multi - jet cross-sections in e+ e- annihilation”. In: *Phys. Lett.* B269 (1991), pp. 432–438. DOI: 10.1016/0370-2693(91)90196-W.

[8] A. Chakraborty et al. “Revisiting Jet Clustering Algorithms for New Higgs Boson Searches in Hadronic Final States”. In: (2020). arXiv: 2008.02499 [hep-ph].

[9] A. Djouadi, J. Kalinowski, and P. M. Zerwas. “Two and three-body decay modes of SUSY Higgs particles”. In: *Z. Phys.* C70 (1996), pp. 435–448. DOI: 10.1007/s002880050121. arXiv: hep-ph/9511342 [hep-ph].

[10] Y. L. Dokshitzer et al. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[11] S. D. Ellis and D. E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[12] A. Hadjighasem et al. “Spectral-clustering approach to Lagrangian vortex detection”. In: *Phys. Rev. E* 93 (6 June 2016), p. 063107. DOI: 10.1103/PhysRevE.93.063107. URL: https://link.aps.org/doi/10.1103/PhysRevE.93.063107.

[13] Hao Li et al. “Strategic Power Infrastructure Defense”. In: *Proceedings of the IEEE* 93.5 (2005), pp. 918–933. DOI: 10.1109/JPROC.2005.847260.

[14] J. Lin. “Divergence measures based on the Shannon entropy”. In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 145–151. DOI: 10.1109/18.61115.

[15] U. von Luxburg. *A Tutorial on Spectral Clustering*. 2007. arXiv: 0711.0189 [cs.DS].

[16] S. Moretti and W. J. Stirling. “Contributions of below threshold decays to MSSM Higgs branching ratios”. In: *Phys. Lett.* B347 (1995). [Erratum: Phys. Lett.B366,451(1996)], pp. 291–299. DOI: 10.1016/0370-2693(95)00088-3,10.1016/0370-2693(95)01477-2. arXiv: hep-ph/9412209 [hep-ph].

[17] S. Moretti, L. Lonnblad, and T. Sjostrand. “New and old jet clustering algorithms for electron - positron events”. In: *JHEP* 08 (1998), p. 001. DOI: 10.1088/1126-6708/1998/08/001. arXiv: hep-ph/9804296 [hep-ph].

[18] A. Y. Ng, M. I. Jordan, and Y. Weiss. “On Spectral Clustering: Analysis and an Algorithm”. In: *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. NIPS’01. Vancouver, British Columbia, Canada: MIT Press, 2001, pp. 849–856.

[19] G. P. Salam and G. Soyez. “A Practical Seedless Infrared-Safe Cone jet algorithm”. In: *JHEP* 05 (2007), p. 086. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph].

[20] R. J. Sánchez-García et al. “Hierarchical Spectral Clustering of Power Grids”. In: *IEEE Transactions on Power Systems* 29.5 (2014), pp. 2229–2237.

[21] J. Shi and J. Malik. “Normalized Cuts and Image Segmentation”. In: *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR ’97)*. CVPR ’97. USA: IEEE Computer Society, 1997, p. 731. ISBN: 0818678224.

[22] A. M. Sirunyan et al. “Measurement of the Jet Mass Distribution and Top Quark Mass in Hadronic Decays of Boosted Top Quarks in pp Collisions at s=13 TeV”. In: *Physical Review Letters* 124.20 (May 2020). ISSN: 1079-7114. DOI: 10.1103/physrevlett.124.202001. URL: http://dx.doi.org/10.1103/PhysRevLett.124.202001.

[23] A. Sirunyan et al. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at sqrt s=13 TeV”. In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/06/p06015. URL: http://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[24] T. Sjostrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: http://www.sciencedirect.com/science/article/pii/S0010465515000442.

[25] G. F. Sterman and S. Weinberg. “Jets from Quantum Chromodynamics”. In: *Phys. Rev. Lett.* 39 (1977), p. 1436. DOI: 10.1103/PhysRevLett.39.1436.

[26] M. Wobisch and T. Wengler. “Hadronization corrections to jet cross-sections in deep inelastic scattering”. In: *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999.* 1998, pp. 270–279. arXiv: hep-ph/9907280 [hep-ph].