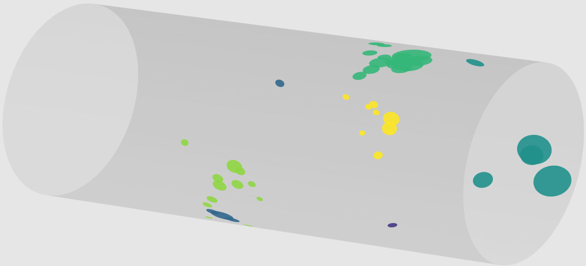
Spectral Clustering for Jet Physics

H.A. Day-Hall*, S. Dasmahapatra, B. Ford,
S. Moretti , C.H. Shepherd-Themistocleous

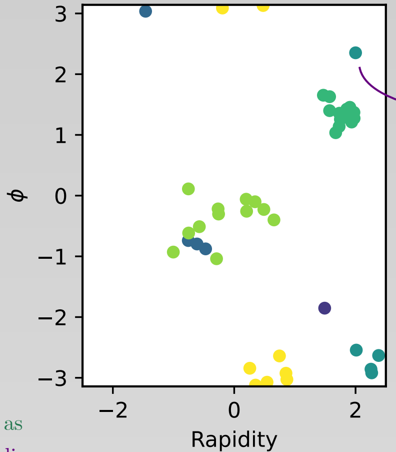
*henrydayhall@protonmail.com



What is a jet?

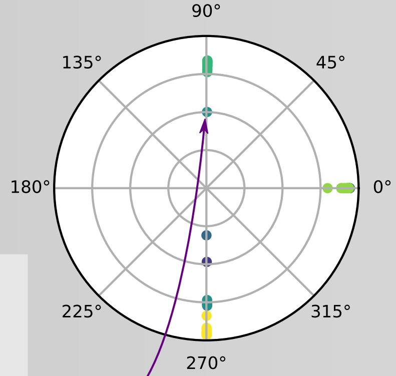
Many particles produced in high energy collisions decay before they can be detected. Clustering the stable particles that are produced is key to reconstructing the particle that decayed. This cluster is called a jet [2].

physical space before first iteration



What is spectral clustering?

Spectral clustering is a Machine Learning (ML) technique for detecting clusters in points with noisy configurations. It requires the construction of a graph Laplacian and calculation of the eigenvectors of the Laplacian to optimise the groupings of the points [3].

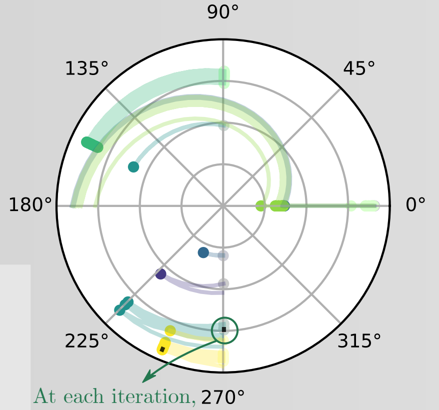


Location of all points/pseudojets are calculated

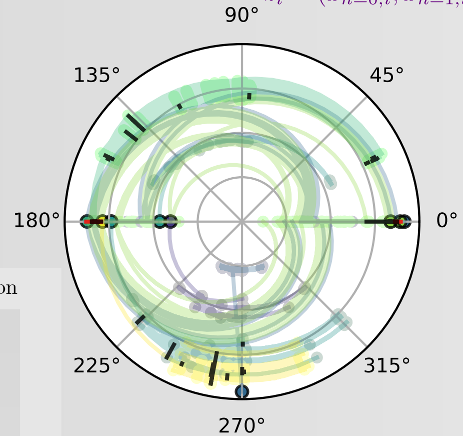
How to form the embedding space

Here is a minimal discription of how to form an embedding space. The full process is discribed in [1].

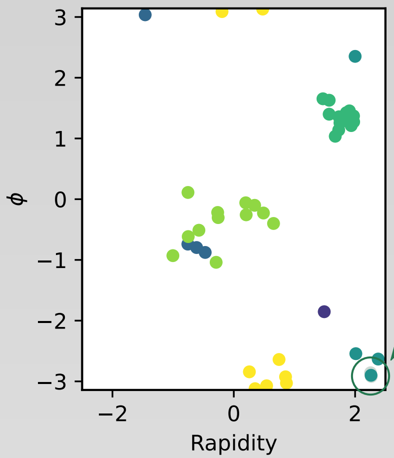
1. Calculate distances between all pairs of points.
$$d_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$
2. Calculate affinities from distances.
$$a_{i,j} = exp(-d_{i,j}^2/\sigma_v)$$
3. Construct Laplacian matrix.
$$D_{i,j} = \delta_{i,j} \sum_k a_{i,k} \quad A_{i,j} = a_{i,j}$$
$$L = D^{-1/2}(D - A)D^{-1/2}$$
4. Find eigenvectors of Laplacian with small eigenvalues.
$$Lx_n = \lambda_n x_n \quad \lambda_n < \lambda_{\text{limit}}$$
5. Define embedding coordinates from eigenvectors.
$$z_i = (x_{n=0,i}, x_{n=1,i} \dots x_{n,i})$$



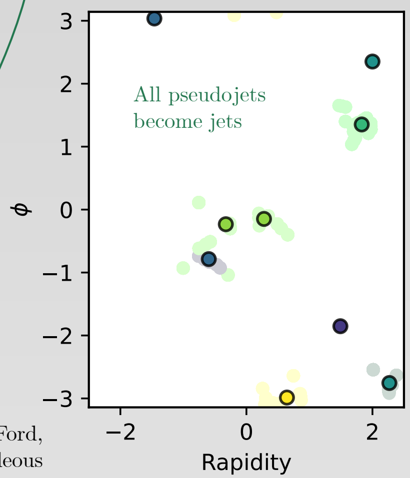
At each iteration, closest two points in embedding space combine



physical space after 2 iterations



physical space after stopping condition



Clustering process

To begin with, all particles are considered pesudojets (candidate jets).

1. Using current pesudojets as points, calculate the embedding space.
2. If there are no dimensions in the embedding space, or the mean distance between points is too high stop clustering.
3. Identify the closest 2 points in embedding space by angle about the origin.
4. Merge the 2 pseudojets corrisponding to the points by combing their 4 momentum (E-scheme recombination).
5. Return to step 1.

When clustering ends, all pseudojets become jets. Jets with less than 2 tracks, or low transverse momentum are removed.

[1] S. Dasmahapatra et al.
Spectral Clustering for Jet Physics. 2021.
arXiv: 2104.01972 [hep-ph].
[2] Gavin P. Salam. “Towards jetography”.
In: The European Physical Journal C 67.3-4
(May 2010), pp. 637–686. ISSN : 1434-6052.
DOI : 10.1140/epjc/s10052-010-1314-6.
[3] Ulrike von Luxburg.
A Tutorial on Spectral Clustering. 2007.
arXiv: 0711.0189 [cs.DS].

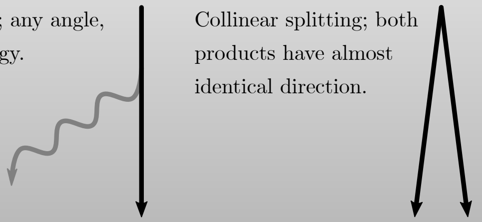
Spectral Clustering for Jet Physics

Infrared and collinear safety

Jet finding algorithms must be insensitive to soft and collinear emissions. That is, the locations and sizes of the jets must not be influenced to particles in this limit.

Soft emission; any angle, very low energy.

Collinear splitting; both products have almost identical direction.



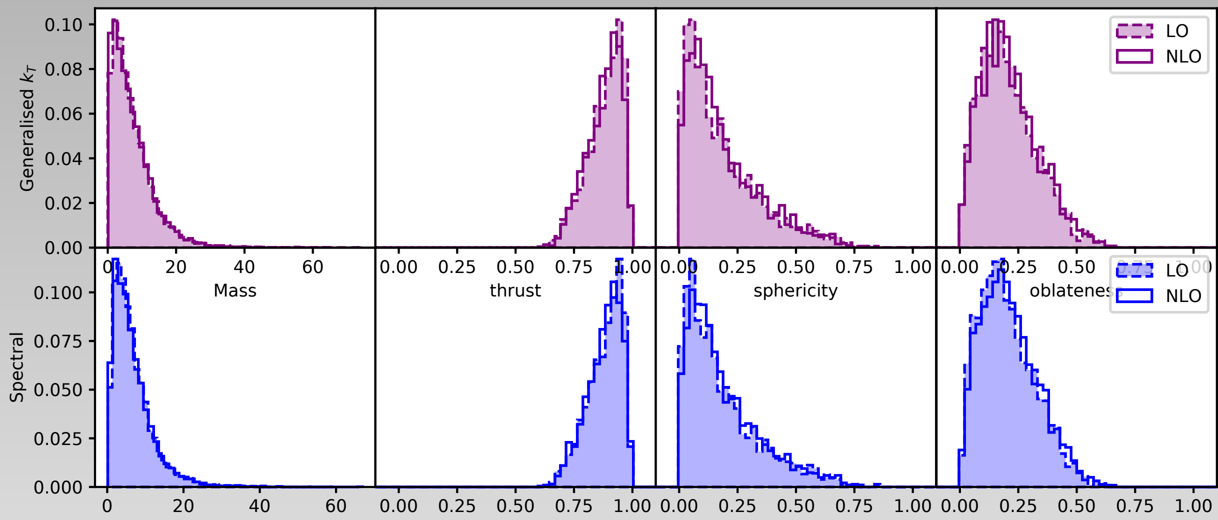
Neither of these may effect the momentum of the jets produced, else the values of the shape variables would be influenced.

Single configuration

Here we check a single parameter configuration (jet radius etc.) for two algorithms. As the shape variables (Mass, Thrust, Sphericity and Oblateness) have the same distributions on LO and NLO data we know both algorithms are insensitive to infrared and collinear radiation (IR-safe).

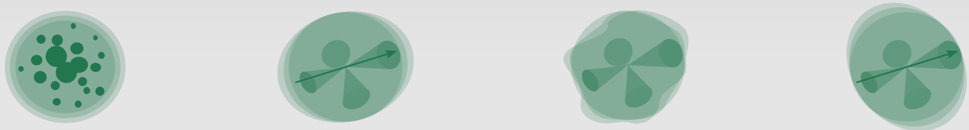
This is known to be IR-safe.

Can we determine if this is IR-safe?



The NLO dataset has soft emissions and collinear splittings, but the LO dataset does not.

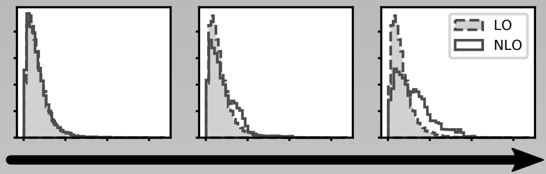
Mass	Thurst	Sphericity	Oblateness
Mass of the total momentum of each jet in all events.	Thrust of the 4 signal jets with highest transverse momentum.	Sphericity of the 4 signal jets with highest transverse momentum.	Oblateness of the 4 signal jets with highest transverse momentum.



Multi-configuration scan

In the top right we compare shape variables for one configuration, but what if we change the parameters? Are the clustering algorithms still IR-safe if the jet radius (or other parameters) change?

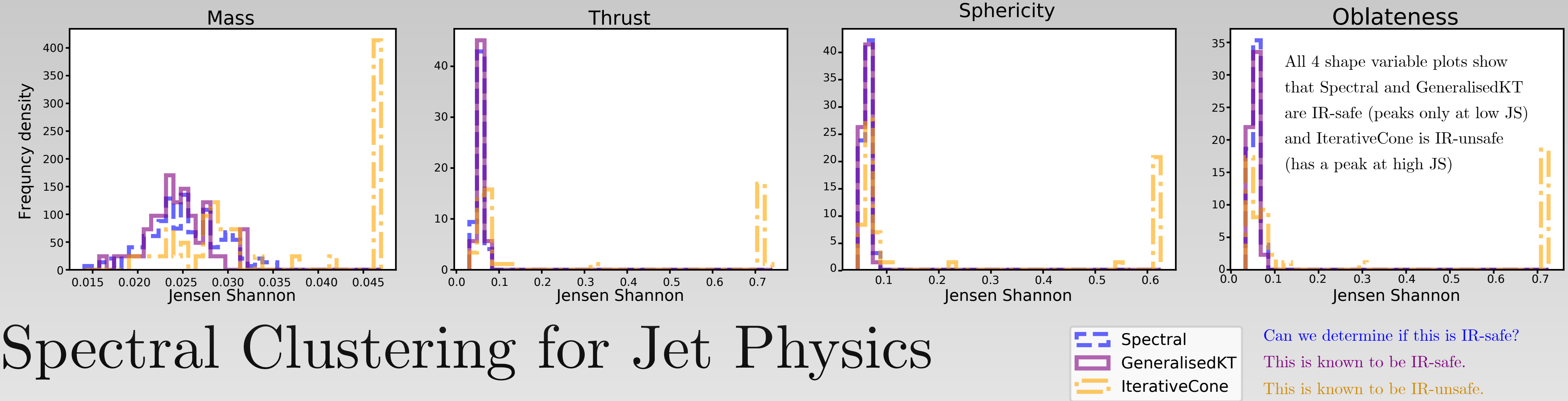
Close LO and NLO distributions, ∴ IR-safe



increasing Jensen Shannon

Different LO and NLO distributions, ∴ IR-unsafe

The Jensen Shannon score is a summary statstic that will let us compare many parameter configurations in one plot.



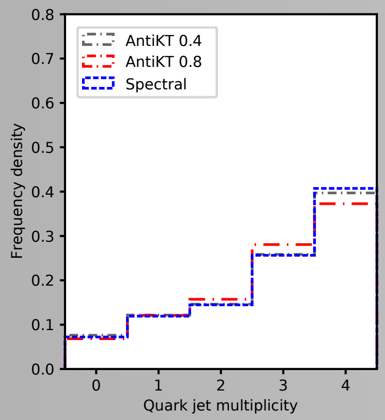
Can we determine if this is IR-safe?

This is known to be IR-safe.

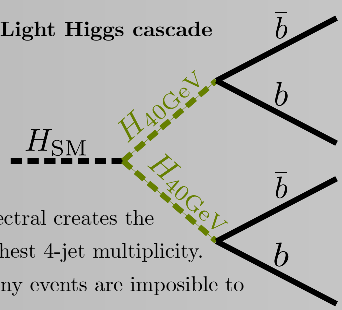
This is known to be IR-unsafe.

Clustering performance Jet formation algorithms are evaluated based on two metrics; jet multiplicity and mass resolution.

Parameter choices for Spectral algorithm optimised for light higgs cascade and held at the same values for other two datasets. Anti-kt algorithms with radius 0.8 and 0.4 are shown on all datasets for comparison.



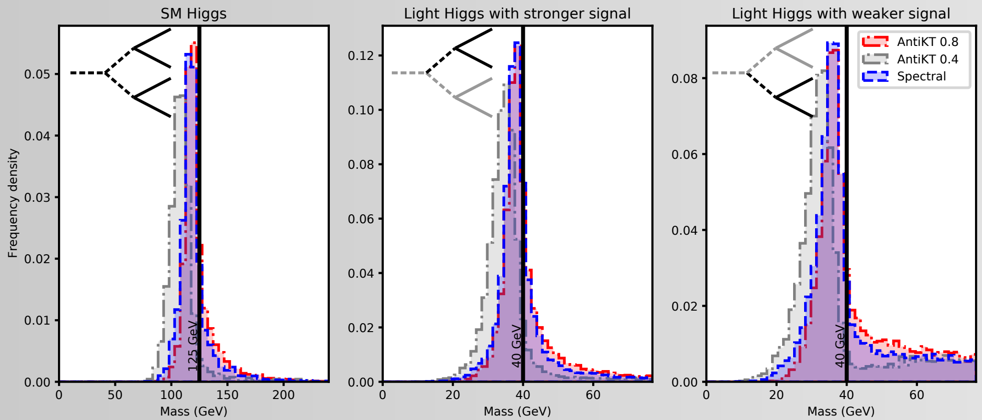
Light Higgs cascade

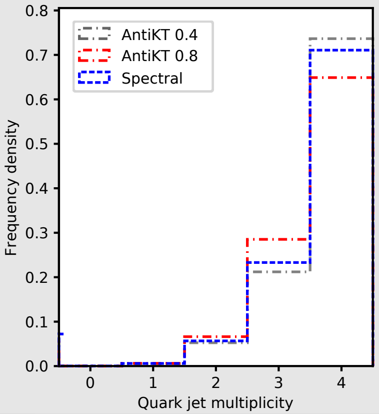


Spectral creates the highest 4-jet multiplicity. Many events are impossible to reconstruct due to low energy.

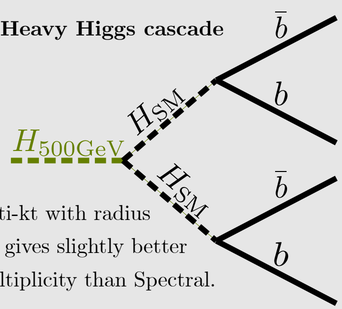
Spectral and anti-kt radius 0.8 both give good narrow mass peaks with the correct location.

Only spectral is producing good mass peaks and multiplicity





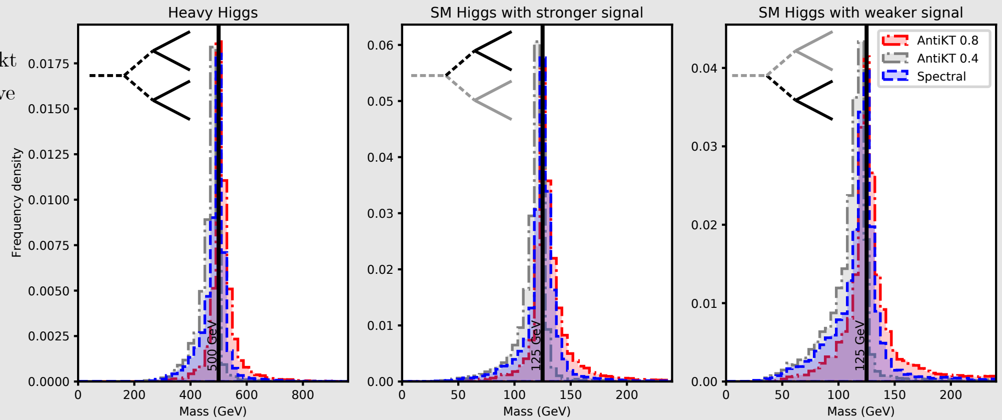
Heavy Higgs cascade

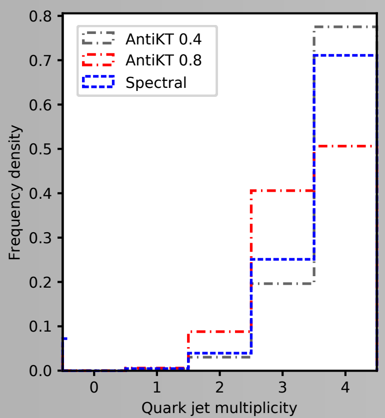


Anti-kt with radius 0.4 gives slightly better multiplicity than Spectral.

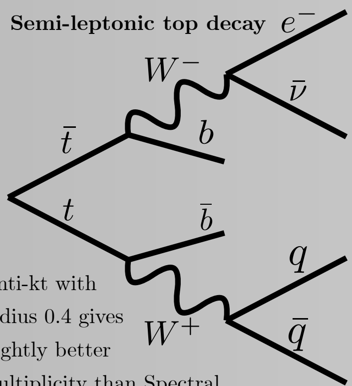
Spectral and anti-kt radius 0.8 both give good narrow mass peaks with the correct location.

Spectral is producing good mass peaks and multiplicity





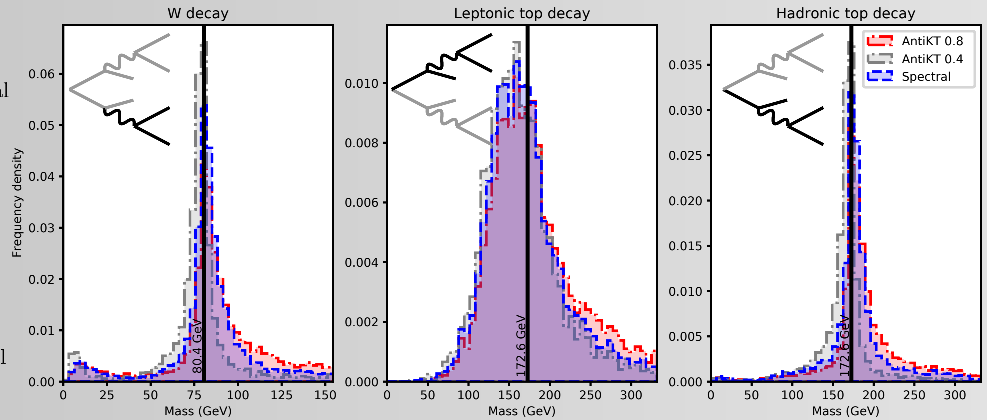
Semi-leptonic top decay



Anti-kt with radius 0.4 gives slightly better multiplicity than Spectral.

Anti-kt radius 0.4 produces the best mass peaks. Spectral is an improvement on anti-kt radius 0.8.

Anti-kt 0.4 is producing good mass peaks and multiplicity, spectral is partially mimicking this.



Spectral Clustering for Jet Physics

Spectral is capable of reconstructing a varierty of events without changing its parameters.