# Jet substructure: a discovery tool

**Deepak Kar**

School of Physics, University of the Witwatersrand, Johannesburg 2050, South Africa

E-mail: `deepak.kar@cern.ch`

**Abstract.** Jets are the collimated bunches of hadrons measured in our detectors, created at high energy particle collisions. As we go to higher energies at the Large Hadron Collider (LHC), Higgs bosons, or yet undiscovered heavy particles are produced with very high energy and the decay products from these "boosted" particles tend to be contained in large radius jets. The internal structure of these jets is exploited to identify the original objects. In this talk, I will motivate the use of substructure techniques for probing new physics at the LHC. I will then discuss the recent experimental results on substructure measurements, including a very new and promising method called *shower deconstruction*.

## 1. Introduction: jets and jet substructure

Most of the interesting collisions at a hadron collider, such as at the Large hadron Collider (LHC) at CERN, involve production of quarks and gluons. When quarks and gluons are produced in a collision, they radiate other quarks or gluons, which in turn radiate again. As they cannot remain as free particles, they form colour-neutral hadrons by combining with other quarks and gluons, which are often created spontaneously from the vacuum. This is observed in the detector as a collimated spray of hadrons, which are termed as jets. There are different algorithms to combine the input objects (i.e. clusters of energy deposits in the calorimeter or simulated particles) in to jets [1]. Along with the specific jet reconstruction algorithm, a jet is also defined by a *size* measure, often termed as radius.

The main aim of the LHC is to discover new particles, which are heavier than already observed Standard Model (SM) particles. Typically, the heavy particles we are interested are not observed directly, but they decay into other particles that are captured by our detectors. Now if this heavy particle is produced with a very high energy, it will be *Lorentz boosted* in the detector center-of-mass frame, resulting in collimated decay products. If the boosted particles decay into a quark anti-quark pair for example, this means the resultant two jets will be largely overlapping, resulting in a single large radius jet in the detector. This discussion is not just relevant for new unseen heavy particles, but top quarks, $W$, $Z$ and Higgs bosons produced with a high enough energy and decaying hadronically can yield such large radius jets.

The idea is then to exploit the internal (sub)structure of these large radius jets to determine their origin, and discriminate the signal against the background. For example, a large radius jet from a hadronically decaying top-quark should have three smaller radius (sub)jets inside, while a large radius jet created from a light quark and gluon should not have such a distinctive characteristic. As the upcoming LHC Run 2 will have more center-of-mass energy compared to Run-1, boosted objects will be more prevalent, and the chances of a discovery might depend on reconstructing them successfully.

Jet grooming and jet tagging are two classes of substructure techniques that have been developed to achieve this, which will be discussed in the subsequent sections.

## 2. Jet grooming

The large-radius jets not only capture the decay products from the heavy boosted object, but also include uncorrelated soft energy deposits, mostly from multiple proton-proton collisions in the same bunch-crossing (termed as *pile-up*), from extra activity associate with the hard scatter (termed *underlying event*) or from initial state radiation. The former is the biggest concern, since the rate of pile-up is going to increase in Run 2.

Three commonly used procedures (collectively called jet grooming) to remove these extra energy deposits, mass-drop filtering [2], trimming [3], and pruning [4] are schematically depicted in Fig. 1, following [5]. They all involve undoing the jet-forming, and using the initial jet constituents.

The mass-drop filtering procedure seeks to isolate concentrations of energy within a jet by identifying relatively symmetric subjets, each with a significantly smaller mass than that of the original jet. The trimming algorithm discards softer constituents with below a certain $p_{\mathrm{T}}$ fraction compared to the jet. The pruning ignores soft or wide-angle radiations.

In Fig. 2, an example result of the jet grooming, as well as the impact of pile-up is shown, following [5]. The left plot shows a comparison of jet mass distribution from simulated signal and background samples before and after grooming. The signal is a simulated $Z'$ particle decaying to quark anti-quark pair produced with HERWIG [6]+JIMMY [7] generator setup, while the background is POWHEG [8, 9, 10]+PYTHIA6 [11] dijet events. It can be seen that after trimming, the $Z'$ mass peak is clearly visible over the background (solid lines), whereas that was not the case before trimming (dotted lines). In the right plot, it can be seen that the ungroomed jet mass increases with the number of pile-up vertices, while after mass-drop filtering (with three different settings), the jet mass has essentially no dependence on pile-up.

## 3. Jet tagging: shower deconstruction

The next step is to come up with observables, distributions of which would be different for large-radius jets coming from signal and background events. Many such observables have been studied with varied amount of success [12]. These observables, and some other dedicated algorithms are used as *taggers*, where a probability of being a signal jet is assigned to each large-large radius jet. A relatively new tagger, called *Shower Deconstruction (SD) [13]*, has yielded very promising results for tagging boosted top quarks [14], and this will be discussed in this section.

The SD algorithm constructs a discriminant, $\chi_{SD}$, optimised to distinguish jets produced in decays of signal particles (S) from jets produced by background processes (B). In the following discussion, the signal process used will be a hadronic top quark decay, and the background process is a jet originating from a single gluon. The SD algorithm is based on the assumption that a final state configuration containing $N$ subjets with four-momenta $\{p\}_N = \{p_1, p_2, ..., p_N\}$ can be generated in many different ways in parton shower approach, and each of these constitutes a possible *shower history*. Then SD calculates the probability that a given shower history was realised in a given event. The shower histories are used to construct the likelihood ratio $\chi_{SD}$ by

$$\chi_{SD}(\{p\}_N) = \frac{P(\{p\}_N|S)}{P(\{p\}_N|B)} \tag{1}$$

where $P(\{p\}_N|S)$ is the probability of obtaining of $\{p\}_N$ given the signal hypothesis, and $P(\{p\}_N|B)$ is then the probability of obtaining $\{p\}_N$ from background jets arising from background processes. $P(\{p\}_N|S)$ and $P(\{p\}_N|B)$ are calculated as the sum of the probabilities for each shower history.

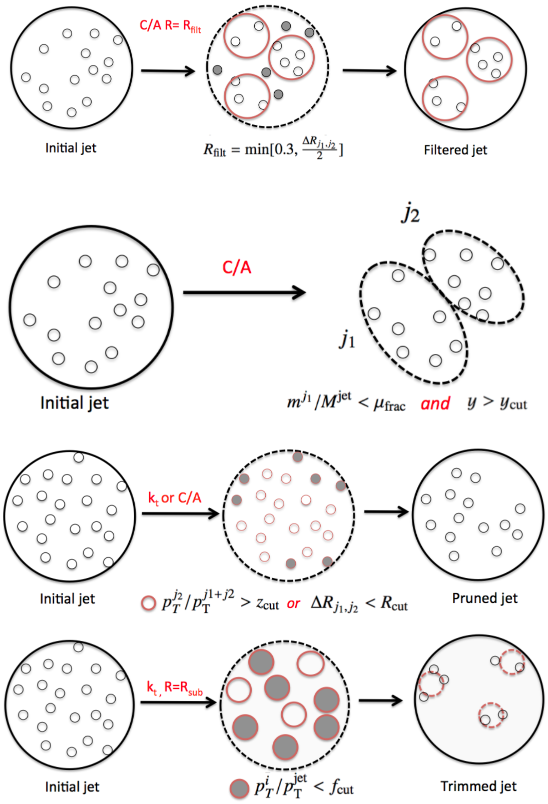

**Figure 1.** Diagrams depicting the the two stages of the mass-drop filtering, jet trimming and jet pruning procedures, from top to bottom.

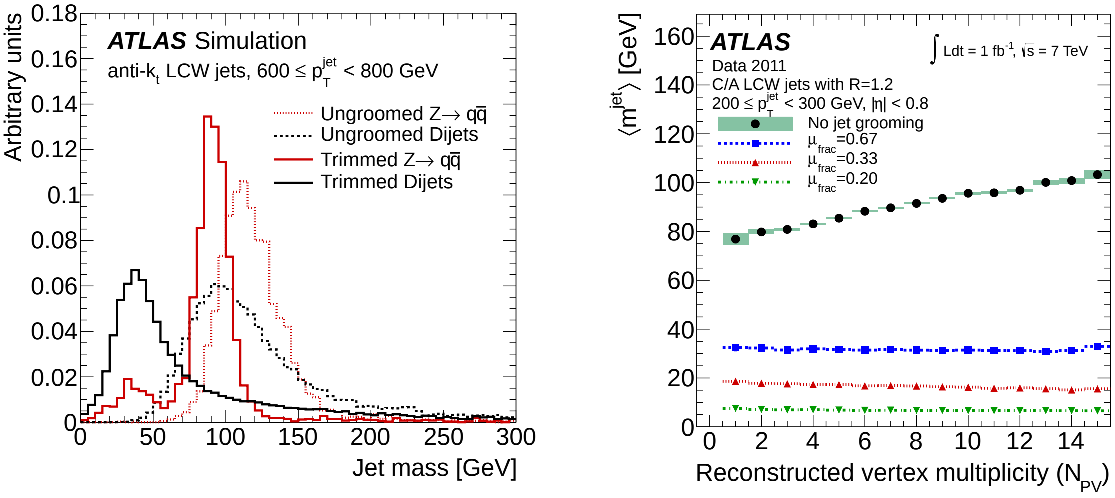

**Figure 2.** Comparison of ungroomed and groomed leading jet mass (left) and stability of groomed jet mass with number of reconstructed vertices (right), using simulated events.

Experimentally the SD is implemented by decomposing the large-radius jet into 0.2 radius subjets, which are used as inputs to the algorithm. The stability of $\chi_{SD}$ against pile-up is an important test of the algorithm. In Fig. 3, top left, it can be seen that indeed $\log \chi_{SD}$ is fairly flat as a function of number of reconstructed vertices. In the top right plot, the mass of the leading composite jet (i.e the jet formed by adding the subjets) is shown, and the peak corresponds to top quark mass peak. These comparisons are done with ATLAS data at a centre-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 14.2 fb$^{-1}$ and signal and background MC samples used in [15].

For the next part, a sample of simulated high-$p_{\mathrm{T}}$ top quarks is used to determine the tagging efficiency. These are generated through a sample of $Z'$ with a mass of 1.75 TeV decaying exclusively to $t\bar{t}$ in the semi-leptonic channel, modelled using PYTHIA8 [16] generator.

The logarithm of $\chi_{SD}$ for the signal and background sample is shown in the bottom left figure, where a clear distinction can be observed. Finally, in the bottom right plot, the Receiver Operating Characteristic (ROC) curve for SD is compared to other commonly used top tagging algorithm [17]. This is obtained by varying the cut on $\log \chi_{SD}$, where background rejection is defined as the reciprocal of the efficiency. The SD algorithm yields the best background rejection over a wide range of signal efficiencies.

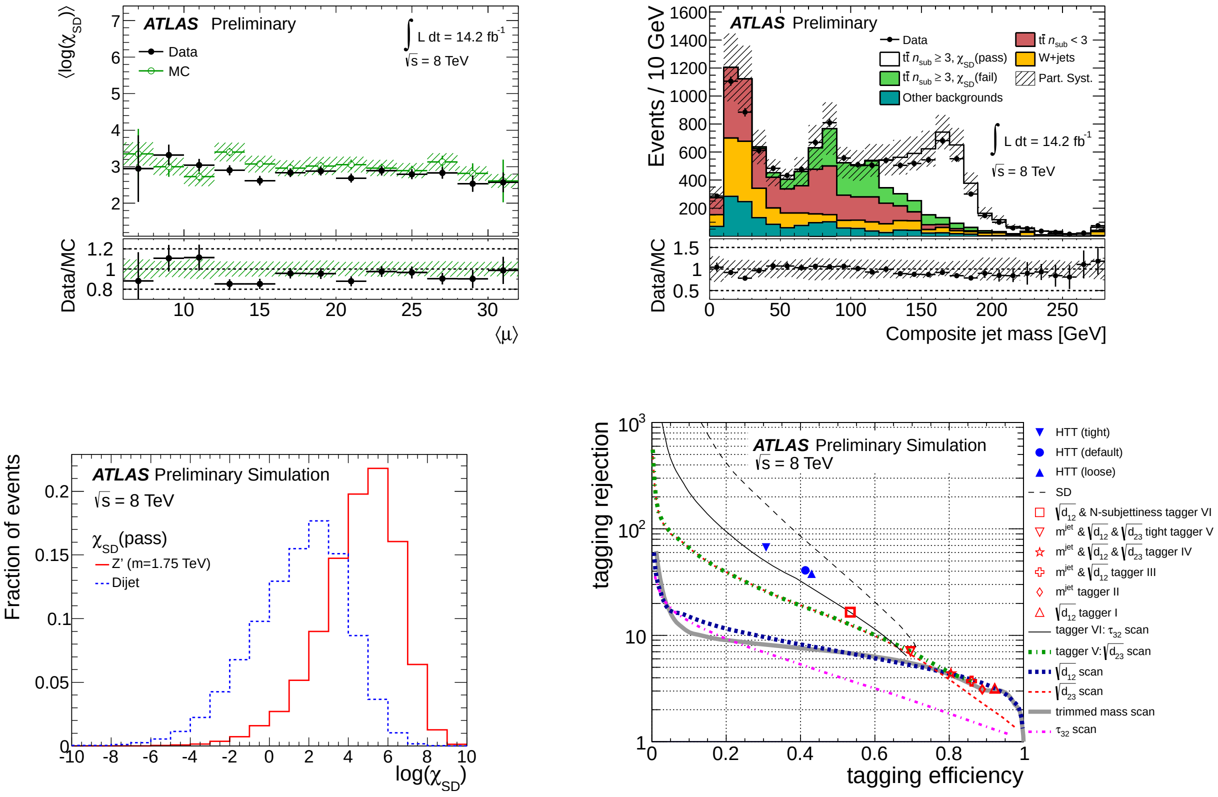

**Figure 3.** Different aspects of SD performance: stability of the $\log \chi_{SD}$ discriminant against pile-up (top left), reconstruction of top quark mass by subjets used in SD (top left), signal and background discrimination using $\log \chi_{SD}$ variable(bottom left) and comparison of SD ROC curve with other taggers (bottom right).

## 4. Summary

Substructure algorithms will be critical in searches for new physics in Run-2. The initial target will be validating the algorithms with Run-2 data, derive experimental uncertainties, and then use them in searches. The Shower Deconstruction algorithm, using the a maximum information approach is one of the promising algorithms, and expected to be used widely.

The performance of this algorithm has been examined in detail for data and MC samples of events predominantly arising from top-quark pair production observed in the lepton plus jets final state. The data were compared to simulation for observables, such as the composite jet mass defined by the mass of sum all of the subjet four-vectors considered by the SD algorithm, and the $\chi_{SD}$ observable. Satisfactory agreement was found between data and simulation as well as stable performance as a function of the pile-up conditions. The expected performance of the SD algorithm and of other top-tagging and substructure techniques has been estimated using samples of simulated high-$p_{\mathrm{T}}$ top quarks from $Z'$ with a mass of 1.75 TeV as the signal and dijets as the background. For this scenario, the SD algorithm shows the best light quark and gluon jet background rejection over a wide range of top-jet signal efficiencies, when systematic uncertainties are not considered.

Shower deconstruction can easily be extended to tag boosted Z, W or Higgs bosons. ATLAS collaboration recently reported an excess in diboson mass spectra in hadronic decay channel [18]. SD and other optimised substructure algorithms will be used to probe this excess in Run 2, and confirm a discovery, if that is the case.

## References

[1] Salam G P 2010 *Eur.Phys.J.C* **67** 637–686 (*Preprint* `0906.1833`)

[2] Butterworth J M, Davison A R, Rubin M and Salam G P 2008 *Phys.Rev.Lett.* **100** 242001 (*Preprint* `0802.2470`)

[3] Krohn D, Thaler J and Wang L T 2010 *JHEP* **1002** 084 (*Preprint* `0912.1342`)

[4] Ellis S D, Vermilion C K and Walsh J R 2010 *Phys.Rev.D* **81** 094023 (*Preprint* `0912.0033`)

[5] ATLAS Collaboration 2013 *JHEP* **1309** 076 (*Preprint* `1306.4945`)

[6] Corcella G *et al.* 2012 Herwig 6.5 release note (*Preprint* `hep-ph/0210213`)

[7] Butterworth J M, Forshaw J R and Seymour M H 1996 *Z. Phys.* **C72** 637–646 (*Preprint* `hep-ph/9601371`)

[8] Alioli S, Nason P, Oleari C and Re E 2010 *JHEP* **06** 043 (*Preprint* `1002.2581`)

[9] Frixione S, Nason P and Oleari C 2007 *JHEP* **11** 070 (*Preprint* `0709.2092`)

[10] Nason P 2004 *JHEP* **11** 040 (*Preprint* `hep-ph/0409146`)

[11] Sjostrand T, Mrenna S and Skands P 2006 *JHEP* **05** 026 (*Preprint* `hep-ph/0603175`)

[12] Altheimer A, Arce A, Asquith L, Backus Mayes J, Bergeaas Kuutmann E *et al.* 2014 *Eur.Phys.J.C* **74** 2792 (*Preprint* `1311.2708`)

[13] Soper D E and Spannowsky M 2013 *Phys.Rev.D* **87** 054012 (*Preprint* `1211.3140`)

[14] ATLAS Collaboration 2014 Performance of shower deconstruction in atlas (*Preprint* `ATLAS-CONF-2014-003`)

[15] ATLAS Collaboration 2013 A search for $t\bar{t}$ resonances in the lepton plus jets final state with atlas using 14 fb$^{-1}$ of pp collisions at $\sqrt{s} = 8$ tev (*Preprint* `ATLAS-CONF-2013-052`)

[16] Sjostrand T, Mrenna S and Skands P 2008 *Comput. Phys. Commun.* **178** 852–867 (*Preprint* `0710.3820`)

[17] ATLAS Collaboration 2013 Performance of boosted top quark identification in 2012 atlas data (*Preprint* `ATLAS-CONF-2013-084`)

[18] ATLAS Collaboration 2015 (*Preprint* `1506.00962`)