Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC

Benjamin Nachman,[a] Pascal Nef,[a] Ariel Schwartzman,[a] Maximilian Swiatlowski[a] and Chaowaroj Wanotayaroj[b]

[a] *SLAC National Accelerator Laboratory, Stanford University,*
  2575 Sand Hill Rd, Menlo Park, CA 94025, U.S.A.
[b] *Center for High Energy Physics, University of Oregon,*
  1371 E. 13th Ave, Eugene, OR 97403, U.S.A.

 E-mail: bnachman@cern.ch, pascal.nef@cern.ch, sch@slac.stanford.edu, swiatlow@slac.stanford.edu, ma.x@cern.ch

ABSTRACT: Jets with a large radius $R \gtrsim 1$ and grooming algorithms are widely used to fully capture the decay products of boosted heavy particles at the Large Hadron Collider (LHC). Unlike most discriminating variables used in such studies, the jet radius is usually not optimized for specific physics scenarios. This is because every jet configuration must be calibrated, insitu, to account for detector response and other experimental effects. One solution to enhance the availability of large-R jet configurations used by the LHC experiments is *jet re-clustering*. Jet re-clustering introduces an intermediate scale $r < R$ at which jets are calibrated and used as the inputs to reconstruct large radius jets. In this paper we systematically study and propose new jet re-clustering configurations and show that re-clustered large radius jets have essentially the same jet mass performance as large radius groomed jets. Jet re-clustering has the benefit that no additional large-R calibration is necessary, allowing the re-clustered large radius parameter to be optimized in the context of specific precision measurements or searches for new physics.

Contents

1 Introduction

The angular separation of decay products for massive particles $\mathcal{P}$, such as W and Z bosons, scales as $2m_\mathcal{P}/p_T^\mathcal{P}$. This suggests that the radius parameter R of jet clustering algorithms aimed at targeting the hadronic decays of $\mathcal{P}$ should be process dependent and scale with the momentum under consideration. However, it is traditionally the case at the Large Hadron Collider (LHC) that only a few choices of R are used for all analyses, which use the best available algorithm (which may not be optimal). This is because every jet configuration, which includes the algorithm, radius, and grooming parameters, must be calibrated to account for unmeasured energy deposits and other experimental effects [1, 2], even though the inputs to jet clustering (topological clusters for ATLAS [3, 4] and particle flow objects for CMS [5, 6]) are themselves calibrated. The calibration of inputs provides a partial calibration to the jet, but jet energy and mass scale corrections provide a *full* calibration by also correcting particles that were missed, merged, or below noise thresholds, energy loss in un-instrumented regions of the calorimeter, and additionally takes into account correlations between particles. The dependence on these additional calibrations thus makes it desirable to reconsider the current jet clustering paradigm in favor of a modular structure that allows for a much broader class of algorithms and radius parameters to be selected by analyses.

One solution is to introduce a new angular scale $r < R$, such that jets of radius r can be the inputs to the clustering algorithm of large radius R jets.[1] If chosen appropriately,

---

[1]Similar ideas have been proposed in the past such as variable R jets [12]. While these methods address the variability of R, they do not address the concerns of individually calibrating many jet collections.

the fully calibrated small radius jets can make the calibration of the re-clustered large radius jets automatic. Furthermore, with no additional calibration needed, any large radius R, any clustering algorithm, and many grooming strategies can be used. Using optimal parameters can, for instance, significantly improve the discovery potential of searches for new physics [7]. In particular, every kinematic region of every analysis for every data-taking condition (e.g. level of pileup - number of additional pp collisions) can optimize these parameters in order to maximize the sensitivity to particular physics scenarios. Another benefit is that the uncertainties on the re-clustered p_T and mass are also automatic consequences of propagating the corresponding uncertainties computed for small radius jets. In this way, the re-clustered jet mass can be viewed as any other kinematic variable, such as di-, tri-, or multi-jet invariant masses that are ubiquitous in measurements and searches for new physics. The uncertainties on re-clustered jets may be higher or lower than those on the corresponding directly constructed large radius jet: a full experimental assessment will be necessary to determine if any precision is lost.

The idea of re-clustering small radius jets is not new. These objects first appeared in an ATLAS search for supersymmetry in the multijet final state [8] and more recently in an ATLAS search for direct stop quark pair production in the all hadronic final state [9]. There are also related techniques which group small radius jets together to form pseudo-jets [10] or mega-jets [11]. However, these analyses still use only a small number of re-grouping techniques which can be further optimized to depend on event kinematics. The purpose of this paper is to describe a systematic comparison of re-clustering jet techniques. Section 2 describes the technical details of re-clustering, including re-clustered grooming which has never been considered before. Section 3 gives the simulation details for the comparison of large radius jets and re-clustered jets presented in section 4. Finally, section 6 provides some concluding remarks.

2   Jet re-clustering

The clustering algorithms at the LHC use sequential recombination. Given distance metrics d_{ij} (between particle 4-vectors) and d_{iB} (between particle and beam 4-vectors), these algorithms recursively combine proto-jets until there are none left. The list of proto-jets is initialized by the set of jet inputs and then at every level of recursion, the algorithms combine particles based on $d_{ij} = \mathrm{argmin}_{i',j'}\{d_{i'j'}, d_{i'B}\}$. If $j \neq B$ then proto-jets i and j are combined into a new proto-jet with a four-vector that is the sum of the four vectors of i and j.[2] If $i = B$, then the proto-jet i is declared a jet and removed from the list of proto-jets.

The most widely used metrics come from the k_t family: $d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n})R_{ij}^2/R^2$, where $R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}$. For $n = 1$ the algorithm is called k_t[3] [14], for $n = 0$ its is called Cambridge-Aachen (C/A) [15, 16], and for $n = -1$ the algorithm is known as anti-k_t [17]. In all three of these algorithms, R is roughly the size of the jet in (y, ϕ) space, though C/A and k_t jets can have very irregular *jet areas* [18].

---

[2]This is a four-vector recombination procedure; other schemes also exist.
[3]See [13] and refs. therein.

The inputs of jet clustering algorithm are typically stable particles (Monte Carlo truth studies), topological clusters (ATLAS), or particle flow objects (CMS). Re-clustered large radius R jets take as input the output of the small radius r jet clustering. Small radius jets have been calibrated with r as small as 0.2 [19] and there are no indications of sizable mis-modelling of close-by effects in the jet response [20]. However, due to falling p_T spectra and a non-negligible jet resolution, the overall closure of the jet calibrations for re-clustered jets needs to be studied by the experimental collaborations. In general, the algorithm used to cluster the small radius jets can be different than the algorithm used for re-clustering the entire event. Figure 1 shows a simple example of an event clustered with anti-k_t $R = 1.0$ and with anti-k_t $R = 1.0$ re-clustered $r = 0.3$ anti-k_t jets. Unlike the inputs of clustering which are e.g. measured in a calorimeter and can be measured and individually calibrated with very low energy, small radius jets can only be reliably fully calibrated for $\gtrsim 15\,\mathrm{GeV}$ [1, 2], where the actual threshold may depend on r. This minimum p_T threshold acts as an effective grooming for the re-clustered jets (RC). This is seen clearly in figure 1, where the blue large radius jet has many constituents far away from the jet axis (which have low p_T) and are not part of the re-clustered jet. One could choose a more aggressive threshold to, for instance, remove the impact of additional pp collisions (i.e. pileup) on the jets. A more dynamic grooming scheme, which we call *re-clustered jet trimming* (RT), in analogy to large radius jet trimming [21], sets the p_T cut on the small radius jets based on the large radius jet p_T (calculated before any small-r jets are removed). Specifically, the RT grooming removes any small radius jet constituent j of a large R re-clustered jet J if $p_T^j < f_{\mathrm{cut}} \times p_T^J$. The parameter f_{cut} can be optimized for a particular kinematic selection and event topology. Other grooming schemes are possible, but beyond our scope.[4]

Due to the increased catchment area of large radius jets over small radius jets, they are more susceptible to contributions from pileup. Just as there are pileup correction techniques for large radius jets and their subjets, one can benefit from pileup corrections to the small radius jet inputs that propagate to re-clustered jets. In particular, one can remove jets from pileup interactions with techniques like JVT [23] or pileup jet identification [24] and can correct the remaining jets with methods like the four-vector jet areas subtraction. Another way to mitigate the impact of pileup is to correct jet constituents before clustering [25–27], which is similar to the $r \to 0$ limit. However, applying a jet constituent pileup correction still requires an overall calibration and an intermediate scale $r \sim 0.1 - 0.5$ is one possibility.

In the growing field of jet substructure, there are many jet observables which depend explicitly on the jet constituents, not just the jet four-vector. These techniques are still applicable for re-clustered jets. Section 5 discusses two approaches to jet substructure in the re-clustering paradigm. In a *top-down* approach, large radius re-clustered jets inherit the constituents of the small radius jets clustered within. Clearly, any constituents that might be part of large radius jets that are not clustered within a small radius jets are not considered under this scheme. However, this removal of radiation also impacts trimmed large radius jets. More details on substructure for trimmed and re-clustered trimmed jets

---

[4]Jet grooming procedures applied to jets-as-inputs have been studied in the past (see for instance ref. [22]); these and other algorithms can be adopted to the re-clustering paradigm.
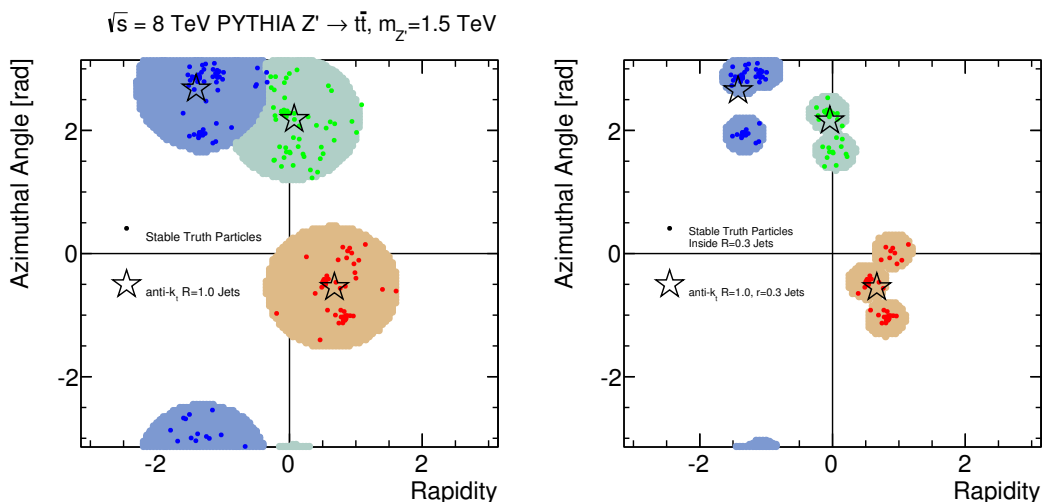
Figure 1. An example event which has been clustered using the anti-k_t $R = 1.0$ (left) and with anti-k_t $R = 1.0$ re-clustered $r = 0.3$ anti-k_t jets (right). The shaded regions show the jet area determined by clustering ghost particles. Only large radius jets with $p_T > 50\,\mathrm{GeV}$ are shown and small radius jets are required to have $p_T > 15\,\mathrm{GeV}$. As with groomed large radius jets, re-clustered jets can have non-circular shapes.

is presented in section 5.1. An alternative *bottom-up* approach to jet substructure is to use the radius r jets directly as the inputs to jet substructure. The advantages and limitations of bottom-up substructure are described in section 5.2.

3 Simulation

Three processes are generated using PYTHIA 8.170 [28, 29] at $\sqrt{s} = 14\,\mathrm{TeV}$ for studying the efficacy of re-clustered jets. Hadronic W boson and top quarks are used for studying hard 2- and 3-prong type jets. To simulate high p_T hadronic W decays, W' bosons are generated which decay exclusively into a W and Z boson which subsequently decay in quarks and leptons, respectively. The p_T scale of the hadronically decaying W is set by the mass of the W' which is tuned to $800\,\mathrm{GeV}$ for this study so that the $p_T^W \lesssim 400\,\mathrm{GeV}$. In this p_T^W range, not all of the decay products of the W are expected to merge into a small radius jet of $r \lesssim 0.4$, but should merge within a cone of $R = 1.0$. A sample enriched in 3-prong type jets is generated with $Z' \to t\bar{t}$, where the Z' mass sets the energy scale of the hadronically decaying top quarks. In this analysis, we use $m_{Z'} = 1.0\,\mathrm{TeV}$, which sets $p_T^t \gtrsim 350\,\mathrm{GeV}$. To study the impact on signal versus background for re-clustering, QCD dijets are generated with a range of $\hat{p}_T$ that is approximately in the same range as the relevant signal process. Pileup is generated by overlaying additional independently generated minimum-bias interactions with each signal event. The value of the number of pileup interactions, n_{PU}, is between LHC Run I conditions, $n_{\mathrm{PU}} = 20$, and the conditions toward the end of the LHC Run II, $n_{\mathrm{PU}} = 80$.

Jet are re-clustered using FASTJET [30] 3.0.3. While the large radius jets can be defined using any set of parameters, the studies in section 4 will use a fixed large jet algorithm: anti-k_t algorithm with $R = 1.0$. Trimmed jet are constructed as follows: the constituents of the large anti-k_t $R = 1.0$ jets are used to reconstruct k_t subjets with a distance parameter $R_{\mathrm{sub}} = 0.3$ and imposing a minimal requirement on the ratio of the subjet p_T divided by the large-R jet p_T (f_{cut}) of 0.1. Here, both transverse momenta are pileup corrected using the median-area method described in ref. [18]. The value of f_{cut} chosen yields approximately optimal resolution of the trimmed jet mass for the studied W' sample — largely independently of the level of pileup considered. Note that this is slightly non-standard: in experimental studies, the f_{cut} parameter is usually imposed on the uncorrected jet p_T, which would require a pileup dependent f_{cut} criterion to optimize the mass resolution. Re-clustering is investigated with a series of schemes for the small radius jets:

- **Algorithms** anti-k_t, k_t, and C/A

- **Radius parameters** $r = 0.2, 0.3$ and 0.4.

- **Grooming** jet p_T cut of 15 GeV for RC and RT with trimming f_{cut} parameters 0.1 and 0.2.

This list is not exhaustive, but encompasses a relevant set of parameters. Radii below $r = 0.2$ are not considered due to experimental limitations from calorimeter granularity and theoretical considerations from non-trivial non-perturbative effects. Small radius jets are required to have $p_T > 15$ GeV.

4 Analysis and results

4.1 Performance of jet mass

The most widely used large radius jet observable is the jet mass, $(\sum_{i \in \mathrm{jet}} E_i)^2 - (\sum_{i \in \mathrm{jet}} \vec{p}_i)^2$, where i runs over the constituents of the jet. To compare the performance of large radius jets with re-clustered jets, we study the performance of the jet mass for the various re-clustering schemes described in section 3. Jet mass performance is quantified by the average jet mass $\langle m \rangle$, a mass distribution standard deviation, σ, and the dependance of these quantities with the amount of pileup. The averages and deviations are computed over a fixed mass range: 60-100 GeV. We also present the efficiency of a $60 < m_{\mathrm{jet}}/\mathrm{GeV} < 100$ mass cut (fraction of events with a mass in this window) as a third figure of merit. First, the small radius jet clustering algorithm is studied, followed by an investigation of the p_T thresholds for RC and RT. A third set of studies compare the performance of the parameter r in RT.

Figure 2 shows the RT jet mass distribution for three small radius jet clustering algorithms with fixed $r = 0.3, f_{\mathrm{cut}} = 0.1$. At very low levels of pileup, there are essentially no differences between building jets from the top down (trimming) or bottom up (re-clustered trimming), as is evident from the fact that $r = 0.3$ k_t jets re-clustered with the $R = 1.0$ anti-k_t algorithm are nearly identical in distribution to the large anti-k_t $R = 1.0$ trimmed jets with k_t subjets. However, with increasing levels of pileup, there are clear differences
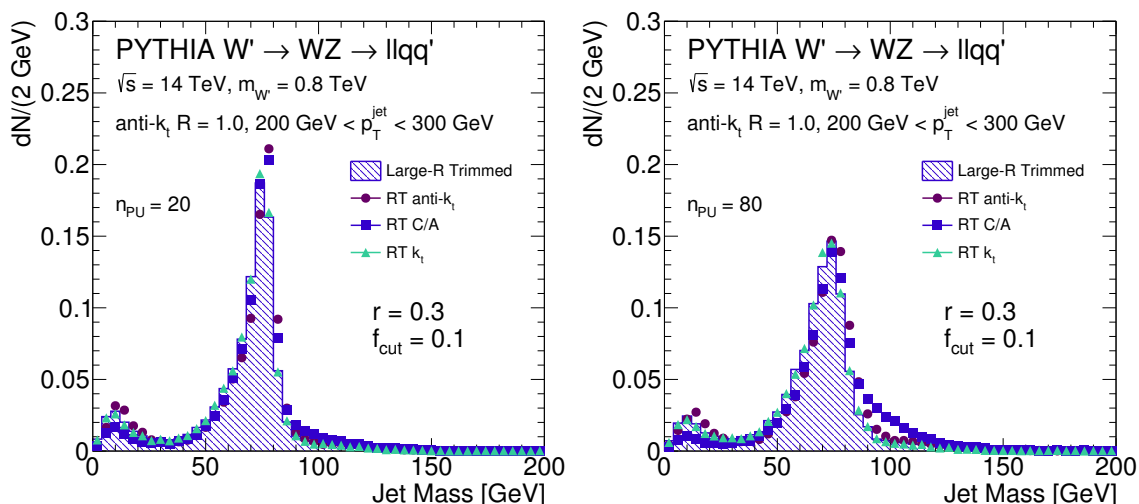
Figure 2. Jet mass distribution for various algorithms using small radius jets at fixed $r = 0.3$ and $f_{\text{cut}} = 0.1$ for $n_{\text{pu}} = 20$ on the left and $n_{\text{pu}} = 80$ on the right, using a sample of $W' \to WZ$ events.

between the three algorithms. For instance, the C/A algorithm has a large high mass tail which is absent in the other algorithms. However, the tradeoff is that C/A has a smaller low mass peak that arises when some of the small radius jets are removed to effectively reconstruct the W jet. Figure 3 is similar to figure 2 but uses a sample of $Z' \to t\bar{t}$ events. Again, especially for low pileup levels, the difference between trimming and the different re-clustering configurations are small. The optimal re-clustering configuration, however, is topology dependent and not necessarily the same for the considered W' and Z' samples.

The discussed features of figure 2 are quantified in figure 4. It is seen that the average mass is higher for C/A jets compared to anti-k_t or k_t, but the efficiency is also highest for C/A jets. Both of these metrics remain stable under increased levels of pileup. Due to its prevalence at the LHC, all subsequent studies will use anti-k_t for the small radius clustering algorithm.

Figures 5 and 6 compare RC with two settings of RT. In the region near the W mass peak, re-clustered trimming with $f_{\text{cut}} = 0.2$ performs the best in terms of the mass distribution standard deviation in the W mass window. However, there is a sizable peak at low mass where too many jets have been cut out by the aggressive trimming parameter. The fixed cut of 15 GeV is too low, especially at very high pileup where the large high mass tail is much bigger for RC than for RT. The re-clustered trimming using anti-k_t with the same f_{cut} as the trimming has very similar performance, though the peak position is slightly higher. Figure 6 shows the performance metrics as a function of n_{pu} for the various grooming schemes. The average mass for RT is very stable, whereas there is a slight slope for RC. The mass resolution for RC is slightly worse than for RT, but the efficiency of RC is better because it avoids the peak at low masses well below the W boson mass.

The re-clustered jet mass distribution for several small radius jet sizes is shown in figure 7 and the performance metrics are quantified in figure 8. For all three considered
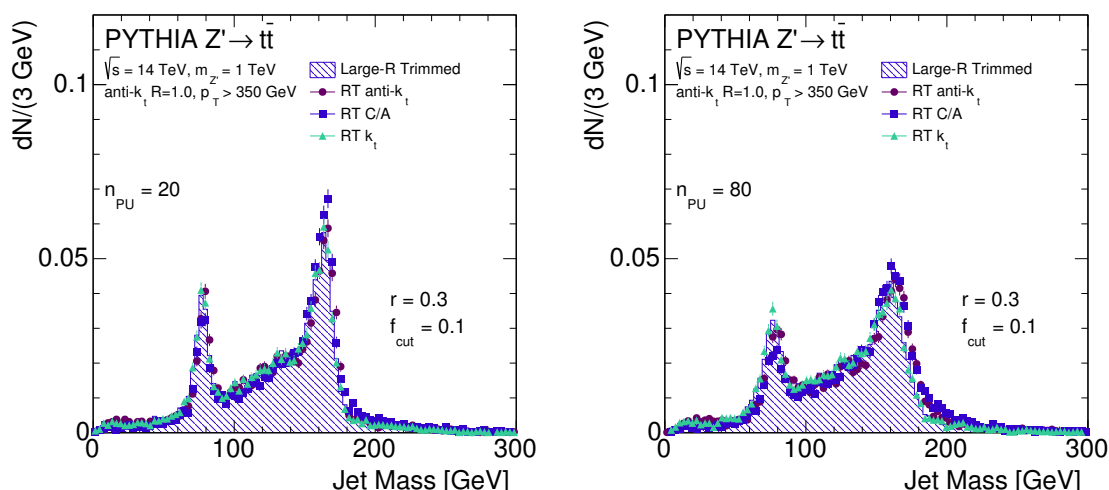
Figure 3. Same as figure 2 but using a sample of $Z' \to t\bar{t}$ events.

values of r, the minimum p_T cut is $15\,\text{GeV}$. In practice, this could be optimized, since smaller radius jets may be calibrated at smaller values of p_T. An alternative approach is to use *iterative re-clustering* by re-clustering $r = 0.2$ into $r' = 0.4$ and then into $R = 1.0$ to further increase the flexibility of the jet algorithms (also this reduces the effective jet area and so the resulting jets would be less susceptible to pileup[5]). The right plot of figure 7 and the top right plot of figure 8 show the $r = 0.2$ as the most peaked.

4.2 Comparisons to QCD

It is important to study the effect of re-clustering not just on signal jets, but also on the most likely background — jets originating from QCD multijet processes. It is well established that various jet grooming techniques increase the separation in jet mass between signal and QCD jets [31, 32]. Figure 9 shows this comparison: while a 4-vector pileup correction alone (left) does not allow for separation between QCD and W jet, both trimming (center) and re-clustered jet trimming (right) allow for the successful discrimination of signal and background using the jet mass. The details of the optimization to improve the signal-to-background in a mass window are best left up to the experimental analyses which use these techniques, but figure 9 indicates that similar performance in QCD rejection is possible with re-clustering as compared to the use of large-R trimmed jets.

5 Re-clustering and jet substructure

Jet substructure techniques have become very sophisticated tools for classifying hadronic final states. There are two natural ways to define jet substructure within the re-clustering paradigm. In a *top-down* approach, re-clustered jets inherit constituents from the radius r

---

[5]If viewed as a uniform noise in the calorimeter, the contribution of pileup to a given jet scales proportionally to its area. There are, however, local fluctuations that complicate this picture.
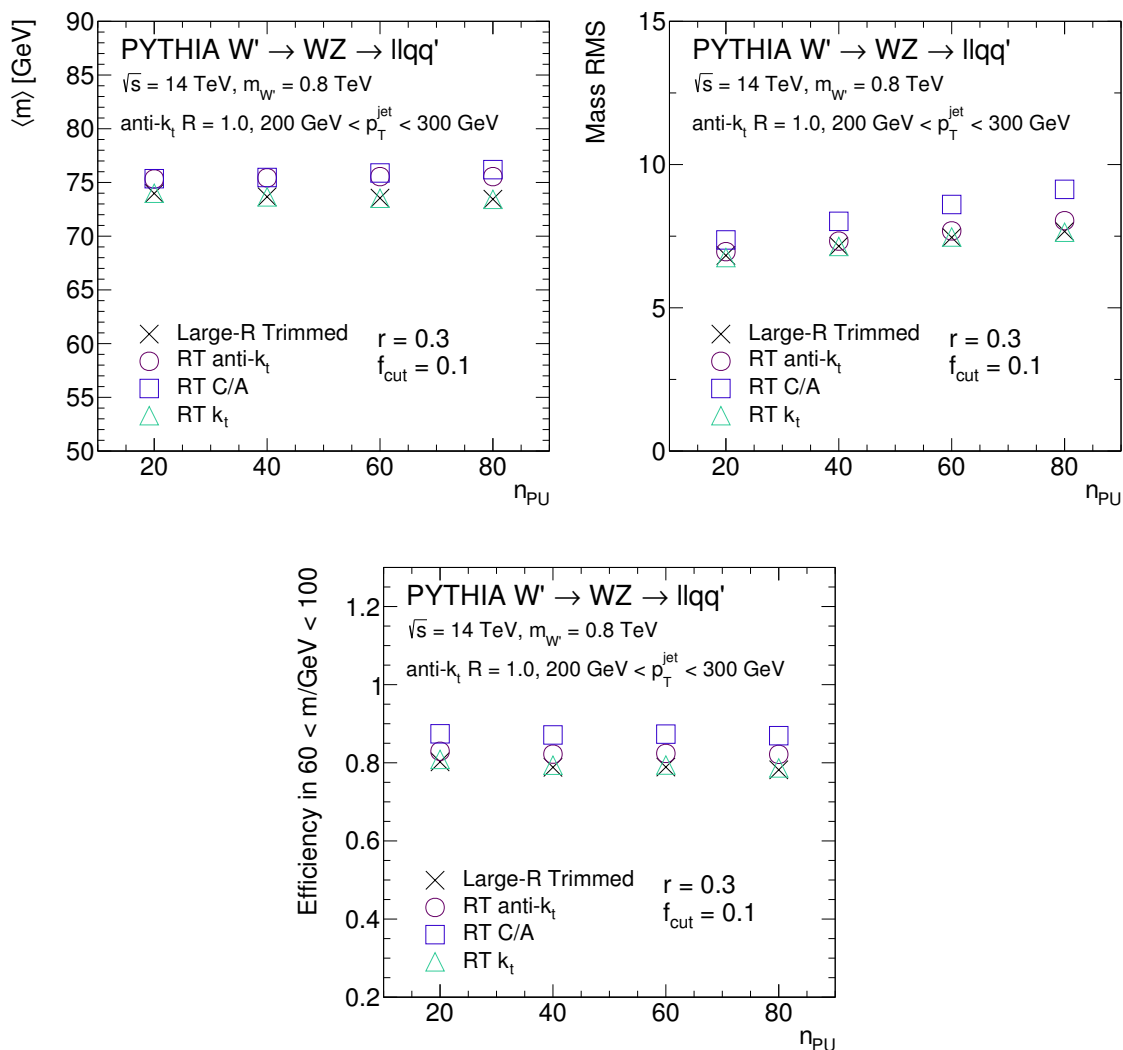
Figure 4. Mean, mass resolution, and mass window efficiency of the mass distribution as a function of the number of additional interaction vertices for various small radius jet clustering algorithms.

jets (section 5.1). In a *bottom-up* approach, the radius r jets are themselves the inputs to calculating jet substructure properties of the reclustered radius R jet (section 5.2).

5.1 Performance of shape variables

In addition to studying the jet mass, in this section we consider one example of a jet substructure variable, the n-subjettiness ratio τ_{21}, and compare QCD to W jets [33]. N-subjettiness moments are defined over a set of N axes,[6] and calculated as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots \Delta R_{N,k}\} \tag{5.1}$$

---

[6]We use the "one-pass" k_t axes optimization technique, which uses an exclusive k_t algorithm to find N axes and then refines them by minimizing the n-subjettiness value.
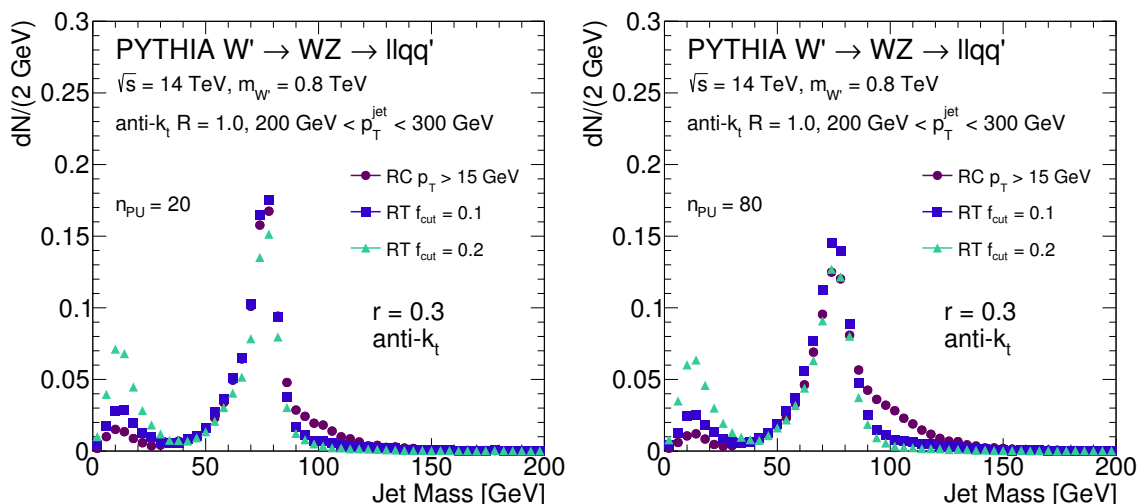
Figure 5. Various re-clustered grooming parameters for anti-k_t $r = 0.3$ jets for $n_{\rm pu} = 20$ on the left and $n_{\rm pu} = 80$ on the right.

with the normalization is defined as:

$$d_0 = \sum_k p_{T,k} R_0 \tag{5.2}$$

where R_0 is the radius of the jet. In practice, analyses use the n-subjettiness ratios:

$$\tau_{ij} = \frac{\tau_i}{\tau_j} \tag{5.3}$$

τ_{21} is often used for the separation of W from QCD jets [34, 35]. This variable measures the compatibility of jets with a 2-prong hypothesis compared to a 1-prong hypothesis, where a low value indicates that the jet likely has a 2-prong structure. When calculating n-subjettiness for re-clustered jets, we use as inputs the constituents of the small jets which have been re-clustered: i.e., we use the truth particles (or in a detector, the particle flow objects or topological clusters) and not the small jets themselves. A pre-selection on the p_T and mass, of $200 < p_T/{\rm GeV} < 300$ and $60 < m/{\rm GeV} < 100$ is applied to all jets considered.

Figure 10 shows a comparison of τ_{21} calculated using trimmed jets (left) and re-clustered jets (right) for samples with $n_{\rm pu} = 80$. Both jet configurations show similar discrimination. For a quantitative comparison, we vary cuts on τ_{21} and plot the W efficiency versus the QCD rejection (defined as the inverse of QCD efficiency). The results, shown in figure 11, are consistent between trimmed and re-clustered jets indicating that powerful discrimination between signal and background jets using jet substructure information is possible with re-clustered jets.

Some studies have shown [34] that un-groomed jets perform better than trimmed or otherwise groomed jets as inputs for the purposes of calculating substructure variables. Re-clustered jets have the disadvantage of always having some effective level of grooming applied (even without the additional re-clustered trimming layer), so in principle calculating
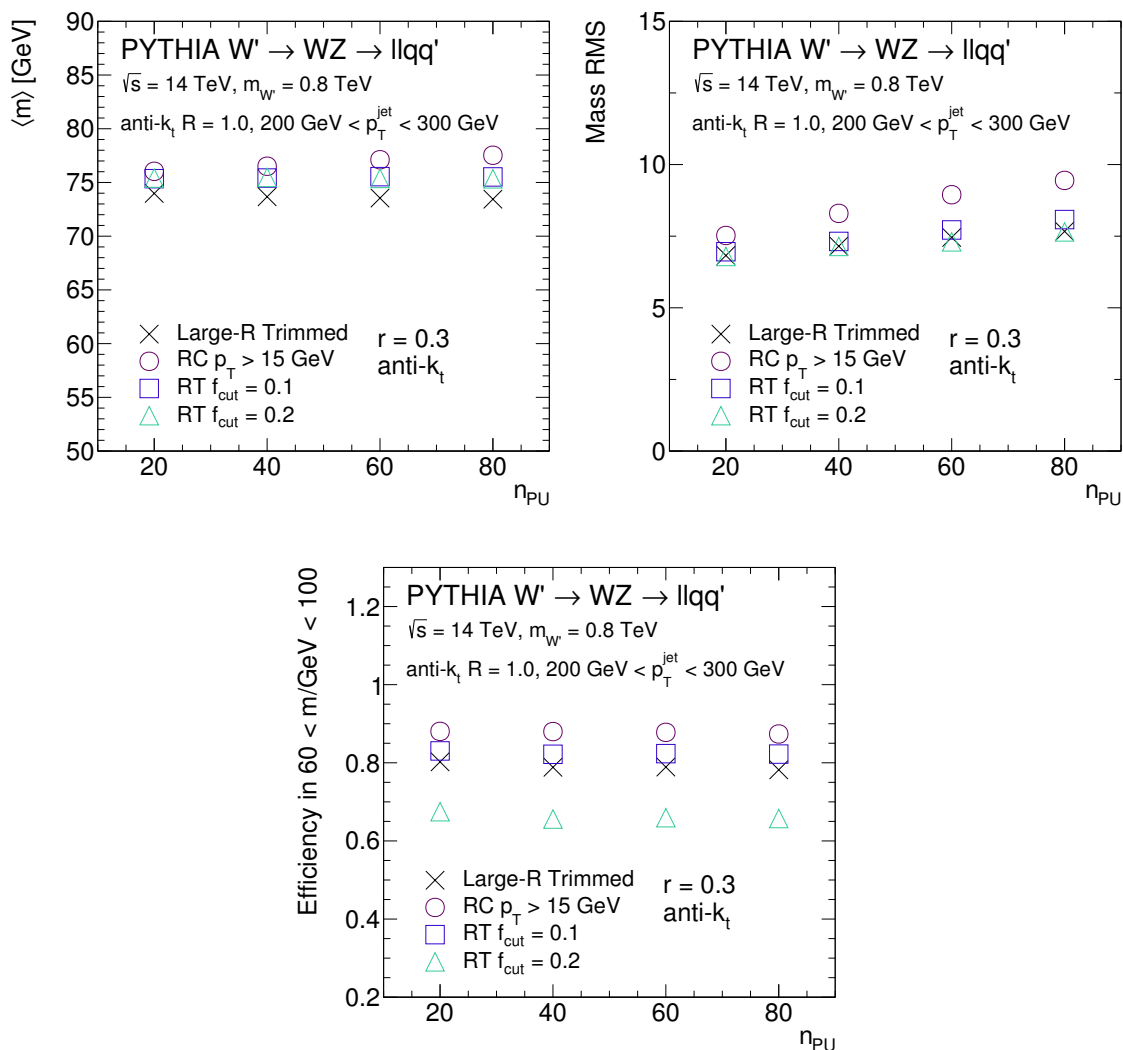
Figure 6. Mean, mass resolution, and mass window efficiency of the mass distribution as a function of the number of additional vertices for various re-clustered jet grooming schemes.

substructure variables without grooming is not possible with re-clustered jets. The event display in figure 1 highlights this issue: substructure variables are sometimes best calculated with large-area jets like those on the left hand plot, but re-clustered jets only ever appear with the reduced-area of the plot on the right. In practice, it is possible to simply associate all constituents closer than the jet parameter R for the purposes of calculating shape variables. The details of such an association are left to studies within the experimental collaborations.

5.2 Bottom-up substructure

An alternative to the assignment of radius r constituents to the radius R jets described in section 5.1 is to use the radius r jets directly. For example, consider the k_t splitting
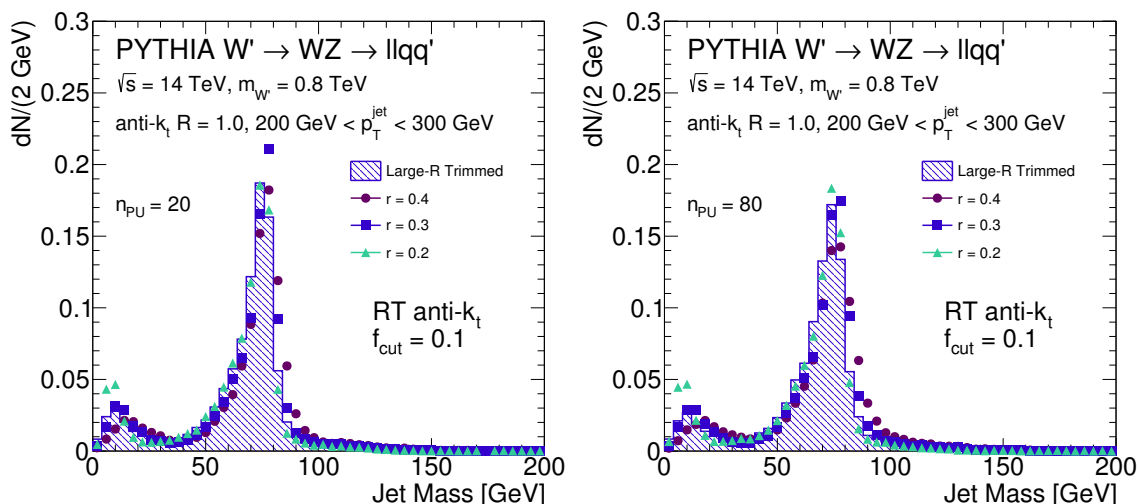
Figure 7. Various small radii for a fixed algorithm of anti-k_t for $n_{\rm pu} = 20$ on the left and $n_{\rm pu} = 80$ on the right.

scale[7] $\sqrt{d_{n,n+1}}$, which is sensitive to hard $(n+1)$-prong structure in a jet. One can use directly the radius r jets inside a radius R re-clustered jet to compute $\sqrt{d_{n,n+1}}$. If there are only two radius r jets, then $\sqrt{d_{12}}$ is simply the k_t distance between the radius r jets. The advantage of this approach is that there is a natural prescription for calibrations and systematic uncertainties. The jet energy scale calibration and its uncertainties directly translate into the calibration of the bottom-up substructure variables. Furthermore, in this approach one knows how the substructure variable calibrations and uncertainties are correlated with the re-clustered jet calibrations and uncertainties. This information is available for the first time with this bottom-up procedure.

Figure 12 compares bottom-up and top-down jet substructure variables in classifying $Z' \to t\bar{t}$ and QCD multijet events. For the chosen parameters, the two techniques have comparable performance. The main drawback of bottom-up substructure is that the relative efficacy depends on p_T (and r). When $r \gtrsim m/p_T$, or equivalently, when there are not many radius r jets inside the radius R jet, the experimental gains from bottom-up substructure are diminished. For instance, if there is only one radius r jet, then $\sqrt{d_{12}} = 0$. Thus, in certain kinematic regimes, bottom-up substructure may provide a powerful alternative to standard methods, but in other regimes a more dedicated analysis is required to understand correlations in calibrations and uncertainties (when jet substructure observables are built from the jet constituents).

---

[7]Computed by re-clustering a jet's constituents using the k_t algorithm and then considering the distance metric of the last n un-clusterings.
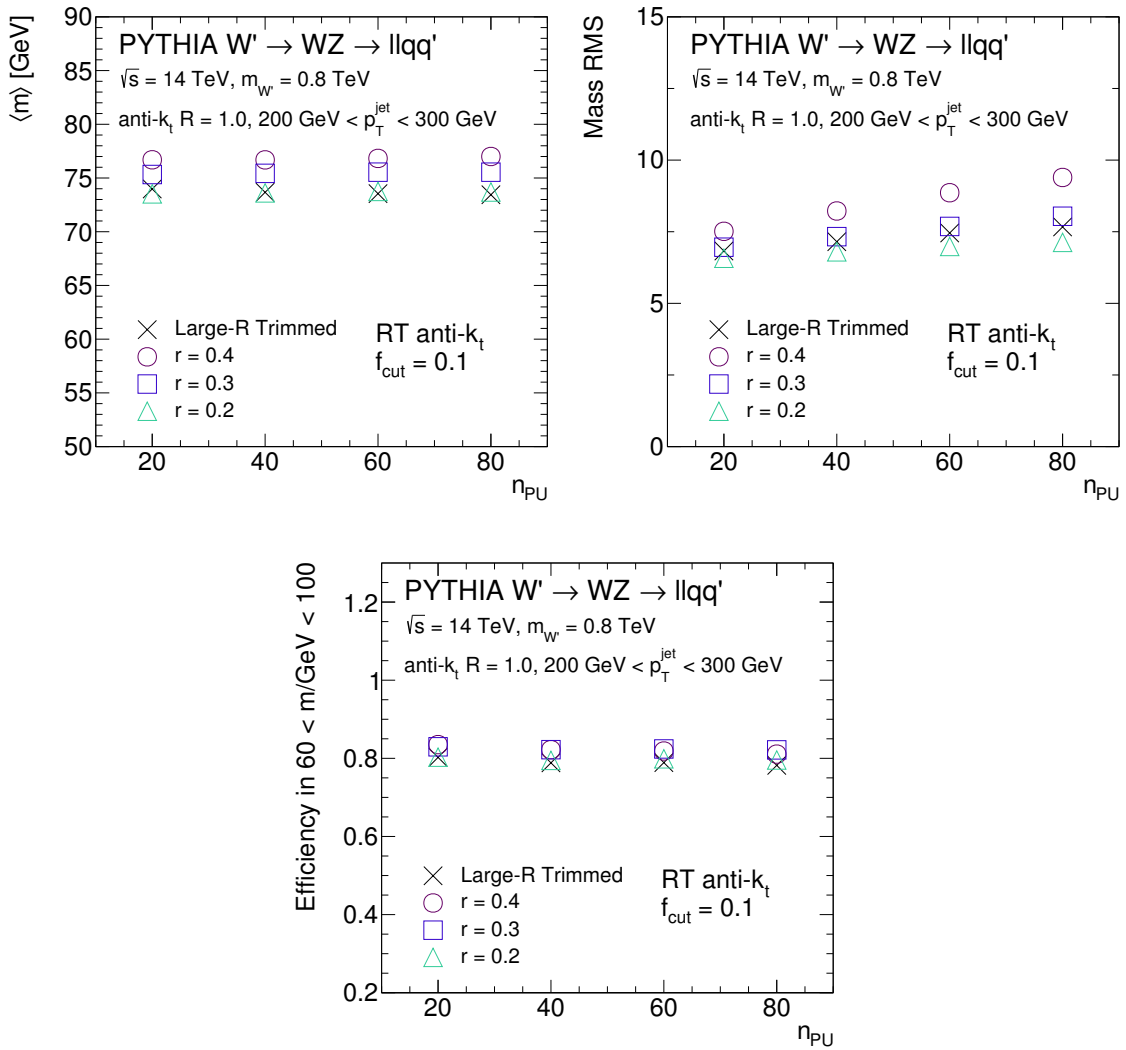
Figure 8. Mean, mass resolution, and mass window efficiency of the mass distribution as a function of the number of additional vertices for various small jet radii.
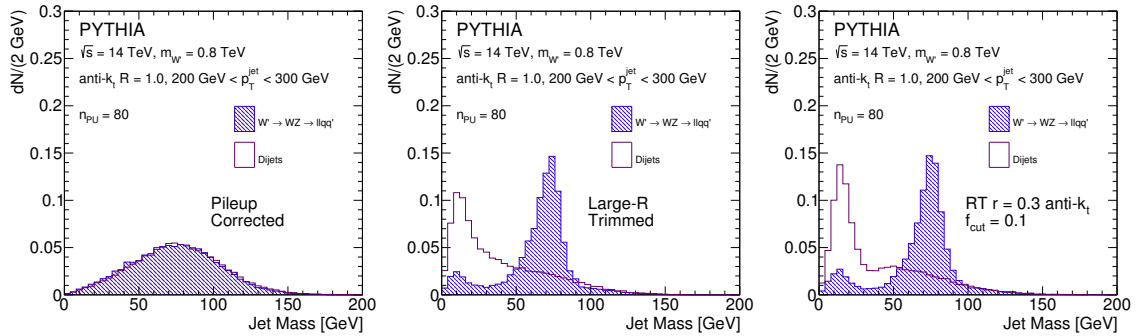

Figure 9. Comparison of jet mass between W jets and QCD jets, for 4-vector pileup corrected jets (left), trimmed jets (center) and re-clustered jets (right). Re-clustered jets use anti-k_t $R = 0.3$ jets as inputs, and apply a trimming cut of $f_{cut} = 0.1$. A pileup level of $NPV = 80$ is used.
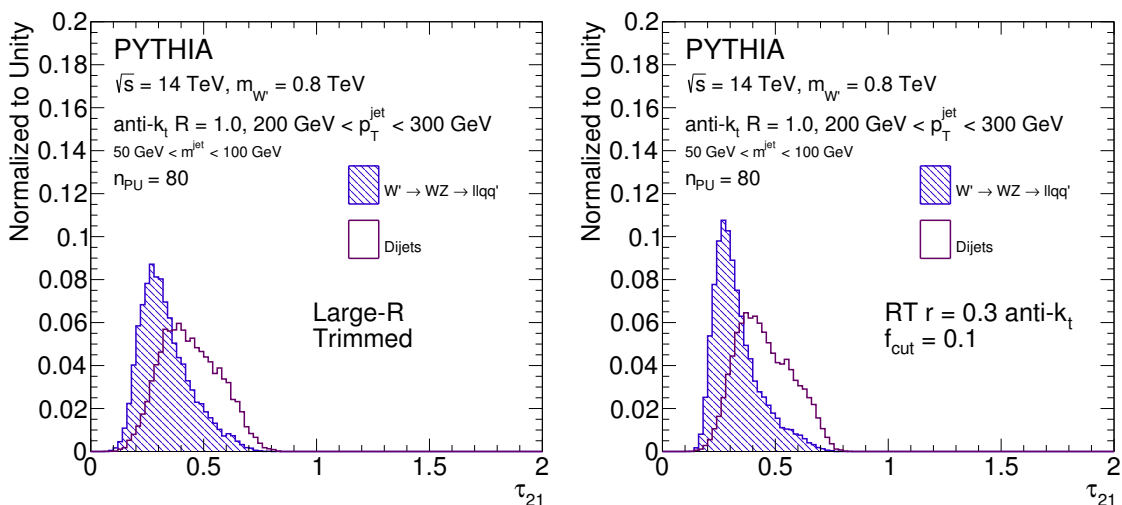
Figure 10. Comparison of the n-subjettiness ratio τ_{21} for W and QCD jets, calculated using trimmed jets (left) and re-clustered jets (right). Re-clustered jets use anti-k_t $R = 0.3$ jets as inputs, and apply a trimming cut of 10%. A pileup level of $NPV = 80$ is used.
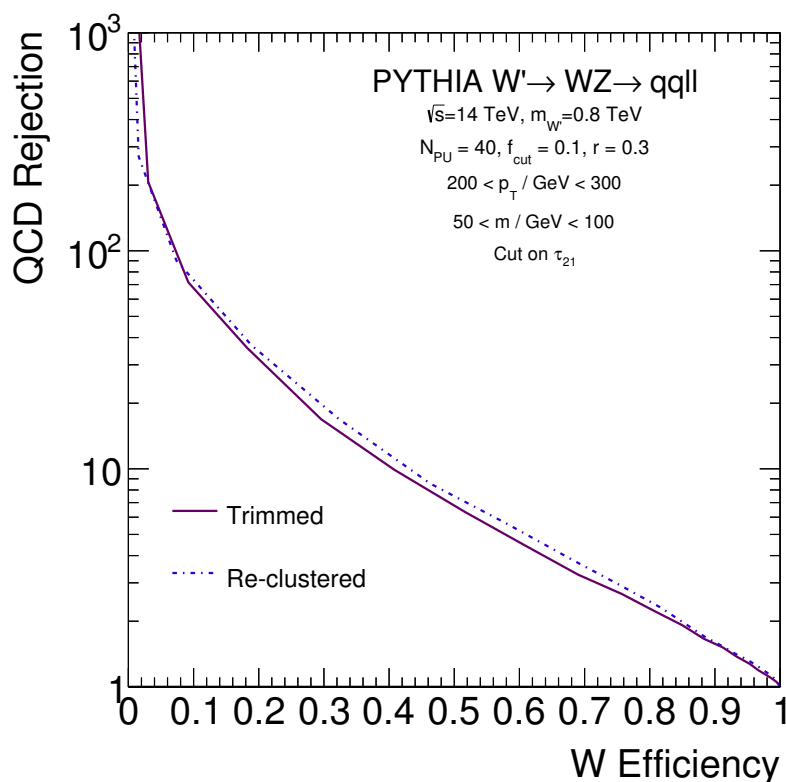

Figure 11. Comparison of the efficacy of cuts on n-subjettiness in separating W from QCD jets. Samples with a pileup of $n_{\mathrm{pu}} = 40$ are shown.
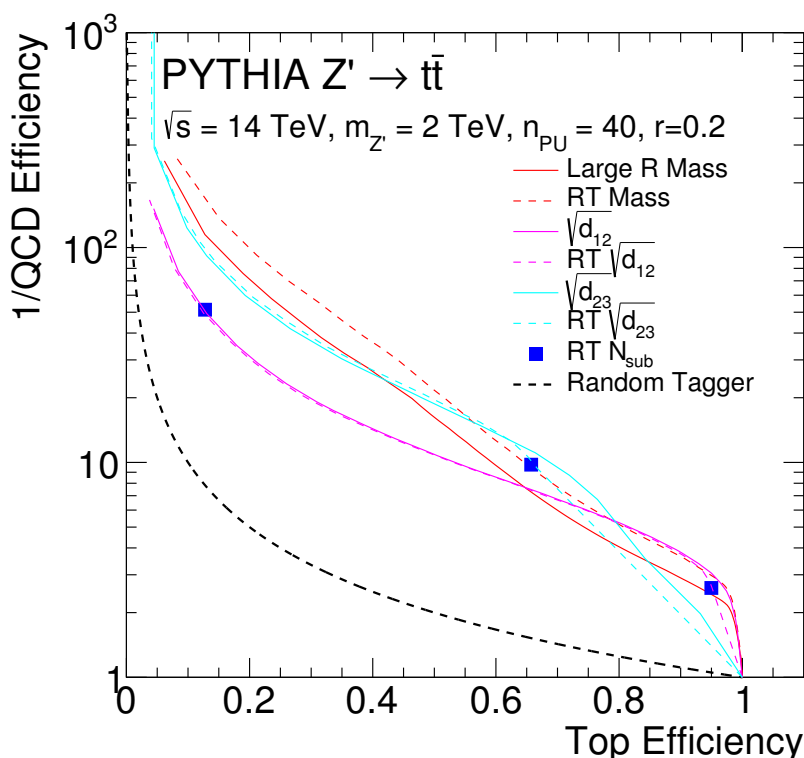
Figure 12. The performance of a bottom-up approach to jet substructure where the radius r jets are inputs to substructure variables. Solid lines show the performance curves for large radius trimmed jets ($R_{\mathrm{sub}} = 0.3$, $f_{\mathrm{cut}} = 0.05$) and the dashed lines show the analogous re-clustered variable. Random tagger denotes a classifier which picks signal and background with equal probability. The variable N_{sub} is the number of re-clustered subjets.

6 Conclusion

Jet re-clustering is a very simple approach for creating large radius jets which emphasizes experimental flexibility by simplifying the commissioning and calibration of jet algorithms. It allows experimentalists to focus on calibrating only one jet algorithm (the small jets used as inputs to larger jets), and has a modular structure that is amenable to optimization per physics process and per event kinematics. Uncertainties and resolutions propagate from the low level to the high level jets, simplifying the determination of mass scales and mass resolutions for large radius jets. Pileup corrections applied to small jets, using both area-based techniques and track/vertex-based approaches, stabilize the large radius jets against pileup. With the addition of re-clustered trimming, performance can be essentially identical to large-R trimmed jets, not only for jet mass but also for shape variables such as n-subjettiness.

Jet clustering — both the choice of algorithm and R-parameter — is an important aspect of analysis optimization in hadronic final states at the LHC. Experimentally, it has been difficult to fully utilize such optimization because of the time and effort involved

in calibrating, validating, and otherwise commissioning each separate collection of jets. Jet re-clustering provides one solution for a modular system that allows every analysis to choose the jet clustering parameters that optimize the sensitivity and accuracy of searches and measurements.

Acknowledgments

References

[1] CMS collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, 2011 *JINST* **6** P11002 [arXiv:1107.4277] [InSPIRE].

[2] ATLAS collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C 75** (2015) 17 [arXiv:1406.0076] [InSPIRE].

[3] W. Lampl et al., *Calorimeter clustering algorithms: description and performance*, ATL-LARG-PUB-2008-002, CERN, Geneva Switzerland (2008).

[4] ATLAS LIQUID ARGON EMEC/HEC collaboration, *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests*, *Nucl. Instrum. Meth.* **A 531** (2004) 481 [physics/0407009] [InSPIRE].

[5] CMS collaboration, *Particle-flow event reconstruction in CMS and performance for jets, taus and MET*, CMS-PAS-PFT-09-001, CERN, Geneva Switzerland (2009).

[6] CMS collaboration, *Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV*, CMS-PAS-PFT-10-002, CERN, Geneva Switzerland (2010).

[7] G.P. Salam, *Towards jetography*, *Eur. Phys. J.* **C 67** (2010) 637 [arXiv:0906.1833] [InSPIRE].

[8] ATLAS collaboration, *Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment*, *JHEP* **10** (2013) 130 [*Erratum ibid.* **01** (2014) 109] [arXiv:1308.1841] [InSPIRE].

[9] ATLAS collaboration, *Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s} = 8\,TeV$ with the ATLAS detector*, *JHEP* **09** (2014) 015 [arXiv:1406.1122] [INSPIRE].

[10] CMS collaboration, *Search for supersymmetry in hadronic final states using M_{T2} with the CMS detector at $\sqrt{s} = 8\,TeV$*, CMS-PAS-SUS-13-019, CERN, Geneva Switzerland (2013).

[11] CMS collaboration, *Search for supersymmetry using razor variables in events with b-jets in pp collisions at $8\,TeV$*, CMS-PAS-SUS-13-004, CERN, Geneva Switzerland (2013).

[12] D. Krohn, J. Thaler and L.-T. Wang, *Jets with variable R*, *JHEP* **06** (2009) 059 [arXiv:0903.0392] [INSPIRE].

[13] 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 [INSPIRE].

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

[15] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323] [INSPIRE].

[16] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, hep-ph/9907280 [INSPIRE].

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

[18] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B 659** (2008) 119 [arXiv:0707.1378] [INSPIRE].

[19] CMS collaboration, *Jet energy corrections for multiple cone sizes webpage*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/MultipleConeSizes14.

[20] ATLAS collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7\,TeV$ with the ATLAS detector*, *Eur. Phys. J.* **C 75** (2015) 17 [arXiv:1406.0076] [INSPIRE].

[21] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, *JHEP* **02** (2010) 084 [arXiv:0912.1342] [INSPIRE].

[22] M. Gouzevitch et al., *Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production*, *JHEP* **07** (2013) 148 [arXiv:1303.6636] [INSPIRE].

[23] ATLAS collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, CERN, Geneva Switzerland (2014).

[24] CMS collaboration, *Pileup jet identification*, CMS-PAS-JME-13-005, CERN, Geneva Switzerland (2013).

[25] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *JHEP* **10** (2014) 059 [arXiv:1407.6013] [INSPIRE].

[26] P. Berta, M. Spousta, D.W. Miller and R. Leitner, *Particle-level pileup subtraction for jets and jet shapes*, *JHEP* **06** (2014) 092 [arXiv:1403.3108] [INSPIRE].

[27] M. Cacciari, G.P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*, arXiv:1407.0408 [INSPIRE].

[28] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A brief introduction to PYTHIA* 8.1, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[29] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA* 6.4 *physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[30] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[31] 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].

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

[33] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].

[34] CMS collaboration, *Identifying hadronically decaying vector bosons merged into a single jet*, CMS-PAS-JME-13-006, CERN, Geneva Switzerland (2013).

[35] ATLAS collaboration, *Performance of boosted W boson identification with the ATLAS detector*, ATL-PHYS-PUB-2014-004, CERN, Geneva Switzerland (2014).