Jet Substructure and Top Tagging

M. Spannowsky
Institute for Particle Physics Phenomenology, Department of Physics, Durham University, DH1 3LE, United Kingdom

Almost all theoretical extensions of the Standard Model predict heavy TeV-scale resonances which have to couple to electroweak-scale resonances, e.g. top quarks or electroweak gauge bosons. Therefore, boosted electroweak-scale resonances with large branching ratios into jets is a highly probable scenario in many processes probing new physics. Here, jet substructure methods can help to disentangle the sought-after signal from the backgrounds. In this brief review we classify scenarios where jet substructure methods can be beneficial for new physics searches at the LHC and discuss the application of the HEPTopTagger in some of these scenarios.

1 Jet substructure in the era of the LHC

The large potential for searches of new electroweak-scale particles by looking inside a fat jet has only been appreciated recently [1,2]. At the LHC with its targetted 14 TeV center of mass energy, particles with masses around the electroweak scale are frequently produced beyond threshold, i.e. boosted transverse to the beam direction. Either because they recoil against other energetic resonances or because they arise from decays of even heavier particles, e.g. Z or KK-gluons. If the resonances transverse momentum is bigger than their mass, their decay products tend to be collimated in the lab frame. Thus, combinatorial problems in the reconstruction of the resonances are ameliorated.

However, at the LHC many sources of hadronic radiation exist. Apart from the decay products of an electroweak-scale resonance, proton-bunch crossings give rise to radiation from the initial state (ISR), the underlying event (UE) and pileup. Initial state radiation results in relatively hard jets. They arise because the incoming partons have to bridge the gap in energy between the proton and the hard process. Underlying event is additional soft QCD activity arising from a given proton-proton interaction surrounding the hard event. It is caused by semi- or non-perturbative interactions between the proton remnants. Finally, pileup denotes multiple proton-proton collisions in one beam crossing.

For an optimal discrimination of a hadronically decaying electroweak resonance from QCD jets, the resonance's decay products have to be disentangled from ISR, UE and pileup.

Sequential jet algorithms [3,4,5], popular for their infrared safety, allow to associate a recombination history to every jet. Therefore, a jet is not only a massive object with a specific cone size and a three-momentum but has a well defined internal structure. Thus, more information is accessible to discriminate the signal from the backgrounds. Over the last few years a plethora of different methods has been proposed to use the internal structure of jets in searches for new physics [2,6]. In general they can be categorized into two classes. On the one hand, methods which extend event shape observables to jet shape observables, making use of the energy flow among the jet's constituents, and on the other hand methods which use internal scales of the recombination history. Often these procedures are combined with so-called jet grooming approaches e.g. Filtering , Pruning or Trimming. At the LHC, the amount of transverse momentum of the underlying event radiation and pileup per unit rapidity can be large [8,9] and their effect on the jet mass depends on the cone size of the fat jet [10]. Grooming procedures are generic prescriptions of how to remove uncorrelated soft radiation from the jet constituents. This goal is achieved by reducing the active area of a jet.

2 Jet grooming methods

Jet grooming methods, like filtering, trimming and pruning, remove soft uncorrelated radiation from a fat jet while retaining final state radiation off the resonance. For QCD jets grooming methods reduce the upper end of the jet mass distribution, whereas for signal events they yield a sharper peak near the true resonance mass $m_j = m_{\text{res}}$. To keep these methods generic it is implicitly assumed that for boosted heavy particles $p_{T_{\text{decay}}} > p_{T_{\text{ISR,UE,PU}}}$.

2.1 Filtering

Filtering, the first proposed jet grooming method, was introduced as part of the so-called BDRS Higgs tagger[1]. Its target application is HW and HZ production with a leptonic decay of the gauge bosons and with the Higgs boson decaying to $b\bar{b}$. A mass drop requirement identifies the vicinity of the Higgs decay products. The procedure called filtering then performs a recombination of the remaining fat jet constituents with a much smaller cone size, R_f. It results in n_f small subjets. This obviously reduces the effective area of the fat jet considered for mass reconstruction and this way tames any QCD effects scaling with R. For the Higgs boson the best mass resolution is achieved by reconstructing the Higgs mass from the $n_f = 3$ hardest filtered subjets. This means we include two b-jets and the hardest wide-angle gluon radiation. Two free parameters, R_f and n_f control the filtering performance.

2.2 Trimming

Trimming [11] targets very similar effects as filtering. In the first step we reconstruct a fat jet which will be heavily impacted by QCD radiation. Its subjets we recombine with a higher resolution R_{trim}, defining a larger number of smaller subjets. These subjets can be separated into two categories: hard and soft. This discrimination is based on the transverse momentum, so hard subjets obey $p_{T,j} > f_{\text{trim}}\Lambda_{\text{trim}}$, where f_{trim} is an adjustable parameter and Λ_{trim} is an intrinsic scale of the fat jet. It can for example be chosen as its jet mass or its transverse momentum. While we discard all soft subjets the recombined hard subjets define a trimmed (fat) jet. Just like filtering this reduces the effective size of the fat jet entering any kind of jet mass measurement. Because Λ_{trim} can be different for each fat jet the trimming procedure is self-adaptive: for a fat jet with large transverse momentum and/or mass the subjets need to have a larger transverse momentum to stay inside the trimmed jet. Just as the filtering procedure, trimming requires two input parameters.

2.3 Pruning

Unlike filtering or trimming, pruning [12] removes underlying event and pileup while building the jet, i.e. as part of the jet algorithm. In a first step it defines a fat jet which can be based on a sequential recombination algorithm. In a second step its constituents are pruned by checking in every recombination step $\min(p_{T,j_1}, p_{T,j_2})/p_{T,j_{1+2}} < z_{\text{prune}}$ and $\Delta R_{j_1,j_2} > R_{\text{prune}}$. If both conditions are met, the merging $j_1, j_2 \to j$ is vetoed. Just as filtering and trimming, pruning depends on two parameters: $z_{\rm prune}$ and $R_{\rm prune}$. $z_{\rm prune}$ ensures that recombined well separated subjets are not very asymmetric in p_T. $R_{\rm prune}$ can be determined on a jet-by-jet basis.

Unlike filtering, pruning and trimming are self-adaptive procedures, applicable to a multi-jet final state in an unbiased resonance search.

It has been shown that pruning, trimming and filtering treat QCD jets differently while yielding a strong correlation in the reconstruction of electroweak scale resonances [13]. Thus, by combining different grooming techniques we can improve the signal-to-background ratio in new physics searches.

3 Phenomenological application of Top Taggers

The reconstruction of boosted hadronically decaying top quarks was one of the first applications of jet substructure methods in searches for new physics [14,15]. In events where boosted top quarks arise from TeV scale resonances top taggers which make use of the substructure of large jets are necessary to discriminate top jets from QCD jets.

Many different approaches to tag boosted top quarks have been proposed [15,16]. It has been shown that they perform similarly on highly boosted top quarks [17]. It is worth noting that it might be possible to combine different top tagging ideas.

One example of a top tagger is the so-called HEPTopTagger (Heidelberg-Eugene-Paris) [18]. The HEPTopTagger is designed to reconstruct top quarks which are only mildly boosted. To capture the decay products of tops with $p_{T,t} \sim 200$ GeV in one fat jet, it is necessary to increase its cone size, e.g. $R = 1.5$. However, increasing the jet area poses two problems for the tagging algorithm. First, subjet combinatorics will increase and it will get more difficult to identify the top decay products. Second, ISR, UE and pileup will become a huge problem, so the HEPTopTagger includes a jet grooming stage.

The tagging algorithm proceeds along the following steps:

1. Un-doing the last clustering of the jet j the mass drop criterion min $m_{j_i} < 0.8\, m_j$ determines if we keep j_1 and j_2. Subjets with $m_{j_i} < 30$ GeV are not considered, which eventually ends the iterative un-clustering.

2. Apply a filtering stage to construct one three-subjet combination with a jet mass within $m_t \pm 25$ GeV.

3. Order these three subjets by p_T. If their jet masses (m_{12}, m_{13}, m_{23}) satisfy one of the following three criteria, accept them as a top candidate:

$$0.2 < \arctan \frac{m_{13}}{m_{12}} < 1.3 \qquad \& \quad R_{\rm min} < \frac{m_{23}}{m_{123}} < R_{\rm max}$$

$$R_{\rm min}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\rm max}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) \quad \& \quad \frac{m_{23}}{m_{123}} > R_{\rm s}$$

$$R_{\rm min}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\rm max}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) \quad \& \quad \frac{m_{23}}{m_{123}} > R_{\rm s}$$

The dimensionless mass windows $R_{\rm min} = 85\% \times m_W/m_t$ and $R_{\rm max} = 115\% \times m_W/m_t$ are tunable and will be optimized by the experimental collaborations.

For the HEPTopTagger the quality of the top quark's momentum reconstruction has been studied in detail. The question if the top tagger really reconstructs all top decay products is surprisingly irrelevant for this test a generic tagger will always be fairly likely to correctly assign the hardest two top decay products, while the softer W decay subjet will contribute little to the reconstructed top momentum. This is the reason why even for moderately boosted tops 95% of the tagged events show a correctly reconstructed direction within $\Delta R = 0.5$; for more than 80% of the tops the momentum is reconstructed within 20% of the Monte Carlo truth (see Fig. 1).

Tagging strategies for medium p_T top quarks can be important for a large variety of applications:

Scalar top partners can ameliorate the top quark's impact on the hierarchy problem of the Higgs, they are among the most anticipated particles to be found at the LHC. The HEPTopTagger was applied [18] to reconstruct the light top squark of the MSSM $\tilde{t}_1$ in a final-state with only jets and missing transverse energy, $pp \to \tilde{t}_1\tilde{t}_1 \to t\bar{t}\chi_1\bar{\chi}_1$. While a reconstruction with standard techniques yields $S/B \sim 1/7$, a subjet analysis in combination with mT2 can result in $S/B \sim 0.88$ and $S/\sqrt{B} \simeq 6$ after 10 fb^{-1}. If one of the tops decays leptonically a leptonic top tagger can be used [19] to separate the neutrinos MET contribution from the neutralinos MET, which allows an effective use of m_{T2}. However, already at 8 TeV and

Recently, CDF [20] and D0 [21] measured an unexpectedly large forward-backward asymmetry of the top quarks. Measurements of this quantity are subtle at the LHC, due to its proton-proton initial state. However, one can define a forward/central charge asymmetry which captures the physics. Unfortunately, for the dominating gg initial state at the LHC there is no asymmetry at all. To enhance the subdominant $q\bar{q}$ and qg production processes it is beneficial to require a large invariant mass of the $t\bar{t}$ system, i.e. require boosted tops. By reconstructing the momentum of the hadronic top and measuring the charge of the second tops lepton, it is possible to count the number of tops and anti-tops in the forward or central region. The forward charge asymmetry we define

$$\mathcal{A}_F(y_0) = \frac{N_t(y_0 < |y| < 2.5) - N_{\bar{t}}(y_0 < |y| < 2.5)}{N_t(y_0 < |y| < 2.5) + N_{\bar{t}}(|y| < y_0 < 2.5))}, \quad (1)$$

for a given rapidity y_0. As shown in the left and middle panels of Fig. 2 this approach allows to measure the Standard Model forward/central charge asymmetry at the LHC [22].

In early ATLAS and CMS reports the $t\bar{t}H$ production channel with subsequent Higgs decay to bottom quarks was one of the major discovery channels for a light Higgs boson. Further studies revealed a very poor signal-to-background ratio of 1/9 [23], making the channel very sensitive to systematic uncertainties which might prevent it from reaching a 5σ significance for any luminosity. However, at high transverse momentum, after reconstructing the boosted, hadronically decaying top quark using the HEPTopTagger as well as the Higgs boson with a modified version of the BDRS method, and requiring 3 b-tags, the signal-to-background ratio can be improved to $\sim 1/2$, while keeping the statistical significance at a similar value to that in Ref. [24], see Fig. 2.

4 Outlook

In this brief review we have given a categorization of new physics scenarios where searches using jet substructure methods can be beneficial over standard search strategies. Any machine probing the multi-TeV scale will produce electroweak scale resonances which will be highly boosted. In this context top tagging is one of the most prominent applications in searches for new physics. The HEPTopTagger is an example of a top tagger applicable in searches for mildly boosted tops. Currently, many different tagging and jet grooming methods are being evaluated on data to test their validity. Present results indicate a huge potential for new physics searches at the LHC in all discussed kinematic scenarios.

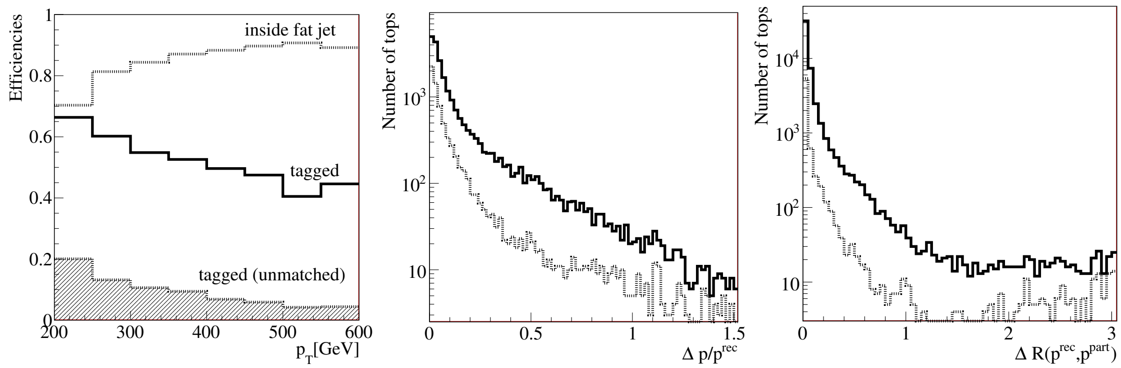

Figure 1: Left: Tagging efficiency of the HEPTopTagger in p_T slices of the fat jet. The dotted line shows the fraction of top quark's where all three decay products on parton level can be found inside the fat jet. The solid black line gives the number of tagged tops. The shaded grey area indicates the ratio of tagged jets where at least one of the subjets is not correctly assigned to the top decay products. Middle: Difference between length of top quark's and reconstructed top jet's three momentum, normalized to the reconstructed top jet's three momentum, i.e. $\Delta p = |p - p^{\rm rec}|$. Right: Angular separation between the reconstructed top jet and the partonic top quark. In the middle and right panels the thin grey line is for $p_{T,j} > 300$ GeV and the solid black line for $p_{T,j} > 200$ GeV.

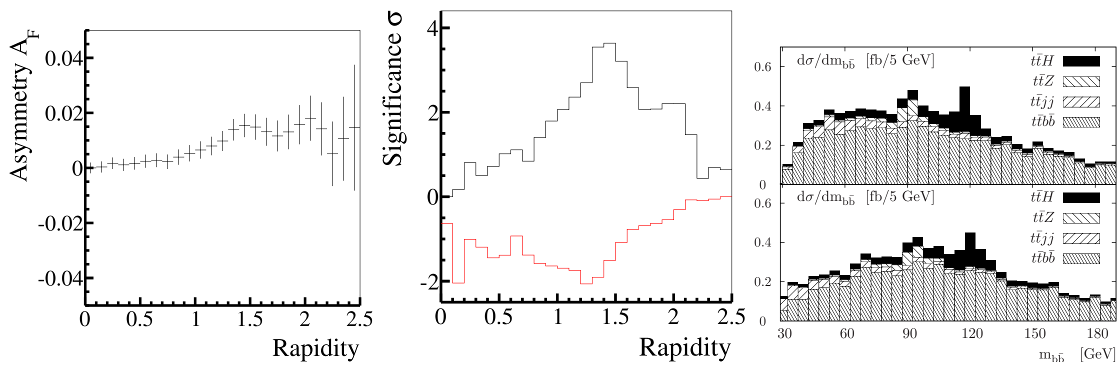

Figure 2: Left: $\mathcal{A}_F$ for a given y_0 according to Eq. (1). Middle: Black curve shows significance to measure the Standard Model $\mathcal{A}_F$. For both panels an integrated luminosity of 25 fb^{-1} at 14 TeV is assumed. Right: Reconstructed Higgs mass, $m_H = 120$ GeV, after event selection cuts and identification of the hadronic top. The upper part shows the reconstruction without underlying event, the lower part with underlying event. The Higgs (and Z boson) peak can be easily disentangled from the shapeless QCD backgrounds.

References

1. J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100** (2008) 242001.
2. A. Abdesselam, *et al.*, Eur. Phys. J. **C71**, 1661 (2011).
3. S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B **406**, 187 (1993); S. D. Ellis and D. E. Soper, Phys. Rev. D **48**, 3160 (1993).
4. Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708**, 001 (1997); M. Wobisch and T. Wengler, arXiv:hep-ph/9907280.
5. M. Cacciari, G. P. Salam, G. Soyez, JHEP **0804**, 063 (2008).
6. L. G. Almeida, R. Alon and M. Spannowsky, arXiv:1110.3684 [hep-ph].
7. ATLAS collaboration, ATL-PHYS-PUB-2009-088.
8. M. Cacciari and G. P. Salam, Phys. Lett. B **659**, 119 (2008).
9. M. Cacciari, G. P. Salam and S. Sapeta, JHEP **1004**, 065 (2010).
10. M. Dasgupta, L. Magnea and G. P. Salam, JHEP **0802**, 055 (2008).
11. D. Krohn, J. Thaler, L. -T. Wang, JHEP **1002**, 084 (2010).
12. S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80**, 051501 (2009); S. D. Ellis,

C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010).
13. D. E. Soper, M. Spannowsky, JHEP **1008**, 029 (2010).
14. K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez and J. Virzi, Phys. Rev. D **77**, 015003 (2008) B. Lillie, L. Randall and L. -T. Wang, JHEP **0709**, 074 (2007).
15. T. Plehn and M. Spannowsky, arXiv:1112.4441 [hep-ph].
16. G. Brooijmans, ATL-PHYS-CONF-2008-008; J. Thaler and L.-T. Wang, JHEP **07** (2008) 092; D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, Phys. Rev. Lett. **101** (2008) 142001; L. G. Almeida, S. J. Lee, G. Perez, I. Sung, J. Virzi, Phys. Rev. **D79**, 074012 (2009).
17. A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, M. Campanelli, B. Chapleau and A. E. Cholakian *et al.*, J. Phys. G G **39**, 063001 (2012) [arXiv:1201.0008 [hep-ph]].
18. T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, JHEP **1010**, 078 (2010).
19. T. Plehn, M. Spannowsky and M. Takeuchi, arXiv:1111.5034 [hep-ph].
20. T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. **D83**, 112003 (2011).
21. V. M. Abazov *et al.* [D0 Collaboration], Phys. Rev. Lett. **100**, 142002 (2008).
22. J. L. Hewett, *et al.*, Phys. Rev. **D84**, 054005 (2011).
23. J. Cammin and M. Schumacher, ATL-PHYS-2003-024.
24. T. Plehn, G. P. Salam, and M. Spannowsky, Phys. Rev. Lett. **104** (2010) 111801.