Eur. Phys. J. C (2016) 76:600
DOI 10.1140/epjc/s10052-016-4443-8

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

A new tagger for hadronically decaying heavy particles at the LHC

T. Lapsien, R. Kogler[a], **J. Haller**

Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany

Received: 17 June 2016 / Accepted: 19 October 2016 / Published online: 3 November 2016

Abstract A new algorithm for the identification of boosted, hadronically decaying, heavy particles at the LHC is presented. The algorithm is based on the known procedure of jet clustering with variable distance parameter R and adapts the jet size to its transverse momentum p_{T}. Subjets are found using a mass jump condition. The resulting algorithm – called Heavy Object Tagger with Variable R (HOTVR) – features little algorithmic complexity and combines jet clustering, subjet finding and rejection of soft clusters in one sequence. While the HOTVR algorithm can be used for the identification of any heavy object decaying hadronically, e.g. W, Z, H, t, or possible new heavy resonances, this paper targets specifically the tagging of boosted top quarks. The studies presented here demonstrate a stable performance of the HOTVR algorithm in a wide range of top quark p_{T}, from low p_{T}, where the decay products can be resolved, to the region of boosted decays at high p_{T}.

1 Introduction

The identification of hadronically decaying heavy standard model (SM) particles (W, Z, H, t) is an important ingredient in an increasing number of SM analyses and searches for new physics at the LHC. For a particle with high energy, the large Lorentz factor leads to decay products which are collimated in the laboratory rest frame and result in a single jet. The task of separating these decays from the vast amount of background from QCD multijet production has been approached with a variety of jet substructure developments in recent years [1–7]. The techniques face the challenge of a stable performance in significantly different kinematic regimes: from the region of low transverse momentum p_{T}, where the decay products can be resolved, to the boosted regime of high p_{T}.

The existing algorithms can be classified into two approaches. The bottom-up approach extrapolates from the resolved into the boosted regime by successively combining small radius jets, similar to an exclusive jet clustering (e.g. the JADE algorithm [8,9]). Modern algorithms have been devised for the task of heavy object identification; examples are the collection of jets in buckets [10,11] or the recently proposed XCone algorithm [12,13]. These methods combine manageable complexity with promising performance, but have not been studied in experimental analyses so far. The larger number of algorithms follows the top–down approach which starts from large radius jets followed by subsequent declustering steps. These algorithms are based on jet clustering with a fixed distance parameter R, where jet grooming methods like filtering [1], pruning [14], trimming [15] or soft drop [16] are used to remove soft radiation and contributions from the underlying event such that substructure observables like the jet mass reflect the hard underlying process. Alternatively the variable R jet algorithm [17] can be used to dynamically reduce the jet distance parameter with increasing p_{T} of the decaying particle. The algorithm was used in studies of new heavy resonances decaying to final states with two and four gluons [15], and also top quark, W and Higgs decays at LHC energies [18], and a similar algorithm at energies of a future hadron collider [19]. Additional substructure information like the k_t splitting scales [20], N-subjettiness [21–23], energy correlation functions [24] or Qjets [25] are often used to further improve the performance of substructure algorithms. Combinations of these methods are used for the tagging of top quarks [26–30], where also more theoretically motivated taggers have been proposed [31,32]. The ATLAS and CMS collaborations have commissioned a number of the techniques mentioned above and studied their behaviour [33–40]. Several top-tagging algorithms have been employed successfully in various searches for new physics [41–69] and in SM top quark measurements [70,71] in LHC analyses with Run 1 data.

[a] e-mail: roman.kogler@physik.uni-hamburg.de

At LHC's Run 2 the production rates of particles with high p_{T} have increased and the importance of boosted analyses is further enhanced. Modifications of existing taggers have been proposed and studied in simulation [72–75]. In most cases, a modest performance improvement is contrasted with a significantly increased algorithmic complexity. A simple but robust algorithm [76] proposed by the ATLAS collaboration has a slightly reduced performance. In addition, recent developments of top–down taggers aim at closing the gap between the resolved and the boosted regime with rather complex algorithmic procedures using several clustering, declustering, mass drop and filtering steps. Examples are the scale invariant tagger [77] and the HEPTopTagger in OptimalR mode [30].

In this work we introduce a new tagger useful in the resolved, the transition and the boosted regime, achieved with only little algorithmic complexity. The tagger is based on the variable R jet algorithm [17], which adapts the jet distance parameter dynamically to the p_{T} of the boosted object. A mass jump condition [78,79] is included in the clustering process, which forms subjets reflecting the dynamics of the underlying hard decay, and enables efficient background suppression. The resulting Heavy Object Tagger with Variable R (HOTVR) accommodates jet clustering, subjet finding and the rejection of soft radiation in one sequence, without the need of declustering and following grooming steps. In this paper we demonstrate the algorithm's properties and characteristics in hadronic top decays and leave studies of the decays of W, Z, H and possible new resonances to future work.

The paper is organised as follows. In Sect. 2 the HOTVR algorithm is described. Its characteristics, free parameters and their influence on the jet and subjet clustering, the collinear and infrared safety and timing performance are discussed in Sect. 3. The algorithm's performance for hadronic top quark decays and a comparison with other commonly used taggers is presented in Sect. 4. A conclusion is given in Sect. 5.

2 The algorithm

The HOTVR algorithm is based on the variable R (VR) jet algorithm [17]. Like all sequential recombination algorithms, it starts with an input list of pseudojets[1] and continues the processing until the input list is empty. The algorithm uses the distance measures d_{ij} and $d_{i\mathrm{B}}$, defined as

$$d_{ij} = \min\left[p_{\mathrm{T},i}^{2n}, p_{\mathrm{T},j}^{2n}\right] \Delta R_{ij}^2, \tag{1}$$

$$d_{i\mathrm{B}} = p_{\mathrm{T},i}^{2n}\, R_{\mathrm{eff}}^2(p_{\mathrm{T},i}), \tag{2}$$

$$R_{\mathrm{eff}}(p_{\mathrm{T}}) = \frac{\rho}{p_{\mathrm{T}}}. \tag{3}$$

The value of d_{ij} can be interpreted as distance between two pseudojets i and j, where $p_{\mathrm{T},i}$ is the transverse momentum of pseudojet i and $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ is the angular distance in rapidity y and azimuth ϕ between the pseudojets i and j. The value of $d_{i\mathrm{B}}$ denotes the distance between pseudojet i and the beam. For a fixed distance parameter of $R_{\mathrm{eff}} = R$ in Eq. (3), the anti-k_t [80], Cambridge/Aachen (CA) [81,82] and k_t [83,84] algorithms are obtained for the choices $n = -1, 0, 1$, respectively. For the HOTVR algorithm $n = 0$ is used, corresponding to CA clustering. However, in the VR algorithm R_{eff} is an effective distance parameter, which scales with $1/p_{\mathrm{T}}$ (cf. Eq. (3)) leading to broader jets at low p_{T} and narrower jets at high p_{T}. The scale ρ determines the slope of R_{eff}. For robustness of the algorithm with respect to experimental effects a minimum and a maximum cut-off for R_{eff} is introduced,

$$R_{\mathrm{eff}} = \begin{cases} R_{\mathrm{min}} & \text{for } \rho/p_{\mathrm{T}} < R_{\mathrm{min}}, \\ R_{\mathrm{max}} & \text{for } \rho/p_{\mathrm{T}} > R_{\mathrm{max}}, \\ \rho/p_{\mathrm{T}} & \text{else.} \end{cases} \tag{4}$$

A known shortcoming of the VR algorithm is the clustering of additional radiation into jets in QCD multijet production, resulting in a higher jet p_{T} on average and an increased rate once a p_{T} selection is applied [17]. The HOTVR algorithm approaches this issue by modifying the jet clustering procedure with a veto based on the invariant mass of the pseudojet pair, inspired by the recently proposed mass jump algorithm [78]. The mass jump veto prevents the recombination of two pseudojets i and j if the combined invariant mass m_{ij} is not large enough,

$$\theta \cdot m_{ij} > \max\left[m_i, m_j\right]. \tag{5}$$

The parameter θ determines the strength of the mass jump veto and can be chosen from the interval [0, 1]. The mass jump criterion (5) is only applied if the mass m_{ij} is larger than a mass threshold μ

$$m_{ij} > \mu. \tag{6}$$

In case a mass jump is found and the p_{T} of the pseudojets i and j fulfil

$$p_{\mathrm{T}i,j} > p_{\mathrm{T,sub}} \tag{7}$$

the pseudojets are combined. The resulting pseudojet enters the next clustering step and the initial pseudojets are stored

[1] We use the notation pseudojet to denote entities entering the jet clustering. These can be partons, stable particles, reconstructed detector objects or combined objects from a previous clustering iteration.

as separate subjets. In case the mass jump criterion is not fulfilled or the pseudojets are softer than $p_{\mathrm{T,sub}}$, the lighter pseudojet or the one too soft is removed from the list. This step reduces the effect of additional activity (soft radiation, underlying event, pile-up) and effectively stabilises the jet mass over a large range of p_{T}.

The full HOTVR algorithm can be summarised as follows.

(1) If the smallest distance parameter is $d_{i\mathrm{B}}$, store the pseudojet i as jet and remove it from the input list of pseudojets.
(2) If the smallest distance parameter is d_{ij} and $m_{ij} \leq \mu$, combine i and j.
(3) If the smallest distance parameter is d_{ij} and $m_{ij} > \mu$, check the mass jump criterion $\theta \cdot m_{ij} > \max[m_i, m_j]$.
   (a) If the mass jump criterion is not fulfilled, compare the masses of the two pseudojets and remove the one with the lower mass from the input list.
   (b) If the mass jump criterion is fulfilled, check the transverse momenta of the subjets i and j.
      (i) If $p_{\mathrm{T},i} < p_{\mathrm{T,sub}}$ or $p_{\mathrm{T},j} < p_{\mathrm{T,sub}}$, remove the respective pseudojet from the input list.
      (ii) Else, combine pseudojets i and j. Store the pseudojets i and j as subjets of the combined pseudojet. In case i or j have already subjets, associate their subjets with the combined pseudojet.
(4) Continue with (1) until the input list of pseudojets is empty.

The algorithm results in jets with an effective size depending on p_{T} and associated subjets. It incorporates jet finding, subjet finding and the rejection of soft radiation in one clustering sequence.

The algorithm is available as plugin to FastJet [85,86] and can be obtained through the FastJet Contribs package [87]. Its implementation is based on the implementations of the mass jump and VR algorithms in the FastJet Contribs packages ClusteringVetoPlugin 1.0.0 and VariableR 1.1.1, respectively. These implementations have been adapted and modified to make the HOTVR software an independent FastJet plugin.

3 Characteristics and properties

Parameters, jet and subjet finding

In total, the algorithm has six parameters, which are listed in Table 1. While the first three parameters steer the VR part of the algorithm, the last three define the mass jump condition. The default values given in the table have been optimised for top quark tagging in pp collisions at $\sqrt{s} = 13$ TeV.

The original VR algorithm is recovered for $\mu \to \infty$. In this case, for $\rho \to 0$ the algorithm is identical to the CA algorithm with a distance parameter of $R = R_{\mathrm{min}}$. Similarly, for $\rho \to \infty$ the CA algorithm is obtained with $R = R_{\mathrm{max}}$. For values of ρ corresponding to the typical scale of an event (m or p_{T} in the range of $\mathcal{O}(100\,\mathrm{GeV})$) jets are clustered with an adaptive distance parameter between R_{min} and R_{max}. Higher values of ρ result in larger jet sizes.

Table 1 Parameters of the HOTVR algorithm. The default values are given for the top-tagging mode

Parameter	Default	Description
R_{min}	0.1	Minimum value of R_{eff}
R_{max}	1.5	Maximum value of R_{eff}
ρ	600 GeV	Slope of R_{eff}
μ	30 GeV	Mass jump threshold
θ	0.7	Mass jump strength
$p_{\mathrm{T,sub}}$	30 GeV	Minimum p_{T} of subjets

The number of subjets found is modified by the mass jump parameters μ, θ and $p_{\mathrm{T,sub}}$. Once the pseudojets become sufficiently heavy due to clustering, the mass jump threshold μ results in a rejection of soft and light pseudojets. For a fixed value of μ, the strength of this jet grooming depends on the parameters θ and $p_{\mathrm{T,sub}}$. For $\theta = 1$ the condition (5) is always fulfilled and no pseudojets are rejected (equivalent to the case $\mu \to \infty$). Conversely, the case of $\theta = 0$ results in a VR jet clustering which stops as soon as a jet mass of μ is reached. The algorithm results in subjets with a maximum mass of μ. Additional jet grooming is obtained by setting $p_{\mathrm{T,sub}} > 0$. This results in subjets with a minimum p_{T} of $p_{\mathrm{T,sub}}$, effectively removing soft radiation and improving the tagging performance at small p_{T} of the heavy object.

The algorithm's behaviour is visualised in Fig. 1 where two example $\mathrm{t\bar{t}}$ events, generated with Pythia 8 [88–90] at low p_{T} (top row, Event 1) and at high p_{T} (bottom row, Event 2), are clustered with the CA algorithm (left column) and with the HOTVR algorithm (right column). The active catchment areas of the hard jets are obtained using ghost particles [91] and are illustrated by the coloured (orange/blue) areas.[2] The impact of the VR part of the algorithm is nicely illustrated by the largely different jet sizes of the two events clustered with the HOTVR algorithm (right column). The grey regions in the right panels were rejected by the mass jump criterion and are not part of the HOTVR jets. This criterion has largest impact in events at low p_{T} as exemplified in Event 1 (top, right). The HOTVR jets together with their subjets reproduce the kinematics of the top decay adequately, both at low and high p_{T}, demonstrating a better adaptation to the decay topology than CA jets. A similar picture is obtained when comparing HOTVR jets to anti-k_t jets.

[2] The exact borders of the jet areas depend slightly on the specific configuration of the ghost particles.

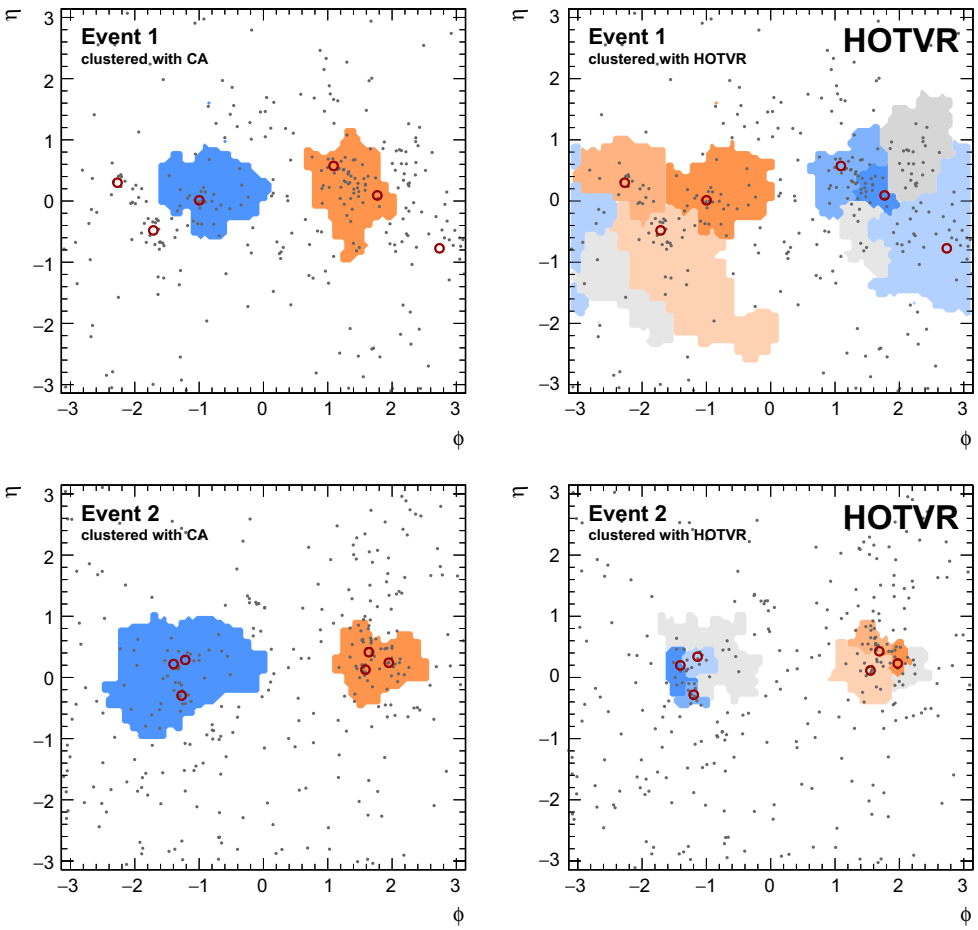

Fig. 1 Two simulated $t\bar{t}$ events clustered with the CA algorithm with distance parameter $R = 0.8$ (*left column*) and with the HOTVR algorithm (*right column*). The *top* quarks have either low p_T (*top row*, Event 1) or high p_T (*bottom row*, Event 2). The two leading jets in the events are shown as coloured areas (*orange*/*blue*). The stable particles, input for the jet finders, are drawn as *grey dots*. The quarks from the *top* quark decay are depicted by *red circles* and are shown for illustration purposes only. In case of the HOTVR algorithm the subjets are *shaded* from *light* to *dark*, corresponding to increasing p_T. The *grey* areas correspond to regions rejected by the mass jump criterion

Collinear and infrared safety

The HOTVR algorithm is infrared and collinear (IRC) safe, except for the unnatural parameter choice of $\mu = 0$. For parameter choices corresponding to the original VR clustering, the HOTVR algorithm is trivially infrared and collinear (IRC) safe [17]. Similarly, for choices of $\mu > 0$ the algorithm is IRC safe, as soft and collinear splittings do not generate mass. This has also been verified in a numerical test, where the stability of the jets as well as subjets found with the HOTVR algorithm was studied with respect to soft radiation and collinear splittings. The algorithm proved to be IRC safe with no events out of 10^6 failing the test [92].

Timing

For timing tests, and throughout this work, the FastJet 3.2.1 [85,86] framework is used, together with FastJet Contribs version 1.024. Starting from FastJet version 3.2, advanced clustering strategies became available which led to substantial speed improvements, especially at high particle multiplicities. For this reason the run time of the algorithm has been studied for different particle multiplicity scenarios, low $\mathcal{O}(50)$, medium $\mathcal{O}(300)$ and high $\mathcal{O}(3000)$. In Table 2 the CPU time of the HOTVR algorithm with default parameters (cf. Table 1) is compared to those of the CA jet algorithm [81,82], the CMS top tagger [26,27], the HEPTopTagger [28,29], the HEPTopTagger in OptimalR mode [30], the VR algorithm [17] as well as the mass jump algorithm [78].

For the various top taggers the CPU time listed includes the time for the underlying jet finding as well as for the top tagger specific processing steps. The developments in FastJet 3.2 result in a much faster runtime of the VR and HOTVR clustering, compared to previous versions (not shown). At low and medium multiplicities, the runtime of the HOTVR algorithm is comparable to that of the other top-tagging algorithms tested. At high multiplicities, it is about a factor four slower than the HEPTopTagger algorithms, but it is still fast

Table 2 CPU time comparison of various algorithms for low, medium and high particle multiplicities. The values are normalised to the CPU time of the CA algorithm with $R = 0.8$

Algorithm	Particle multiplicity		
	$\mathcal{O}(50)$	$\mathcal{O}(300)$	$\mathcal{O}(3000)$
CA algorithm ($R = 0.8$)	1.0	1.0	1.0
CA algorithm ($R = 1.5$)	1.5	1.3	1.3
CMS top tagger ($R = 0.8$)	1.3	1.1	1.1
HEPTopTagger ($R = 1.5$)	1.8	1.3	1.3
OptimalR HEPTopTagger ($R = 0.5$–1.5)	2.4	1.4	1.3
VR clustering ($\rho = 600\,\text{GeV}$)	1.3	1.6	5.8
Mass jump algorithm	1.7	4.2	23.8
HOTVR	1.6	1.7	5.3

enough for practical uses.[3] The original mass jump algorithm has not been updated to employ the new clustering strategies, which leads to a much worse performance at medium and high multiplicities.

4 Physics performance

Studies of the physics performance are carried out using the event generator Pythia 8 [88–90]. A pp $\rightarrow$ $t\bar{t}$ sample is used as signal process, background events are obtained by simulating QCD dijet production in pp collisions. For both samples a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ is used, the multiple parton interaction tune Monash 2013 [93] and the LO NNPDF2.3 QCD+QED [94] PDFs with $\alpha_s(M_Z) = 0.130$ are employed. At this stage no additional pp interactions during a single bunch crossing (pile-up) are simulated.[4]

Throughout this work, jets are clustered using all stable particles from the Pythia 8 output. In some studies, additional jets (labelled parton jets) are obtained using a list of all final state partons[5] as input to the anti-k_t algorithm with distance parameter $R = 0.4$ with a minimum p_T of 100 GeV. In case of $t\bar{t}$ production, the top quark is effectively treated as stable for the purposes of defining the parton jet: after showering the top quarks are added to the parton list, and all partons from the top quark decay are removed. In case a matching between particle and parton jets is employed, the geometrical matching condition $\Delta R < R_\text{eff}$ is used.

Reconstruction of masses and transverse momenta

The key to the tagger's effectiveness is the accurate reconstruction of subjets originating from the top quark decay, achieved by the VR condition and the mass jump criterion. This leads to a stable peak position for the mass of top jets over a large range of jet p_T, as shown in Fig. 2. The jet mass m_jet distribution for jets with two different subjet multiplicity N_sub selections is shown for two ranges in the p_T of the parton jet matched to the particle jet. For $t\bar{t}$ events with $N_\text{sub} \geq 2$ the distributions feature a dominant peak, stable around the top quark mass[6] for fully merged decays, and two smaller peaks at lower masses corresponding to partially merged top quark decays. The requirement of $N_\text{sub} \geq 3$ leads to a depletion of the two secondary peaks, while the peak around the mass of the top quark is hardly affected. At low p_T (left) the top quark peak is wider with a larger tail and is situated on a larger plateau than at high p_T (right) because of contributions from additional radiation which aggregate in the jet due to its large size. While this leads to a larger misidentification rate at low p_T, it results in a non-vanishing efficiency already at top quark transverse momenta as low as 100 GeV. For typical QCD jets a falling distribution is observed. The wide peak at mass values around 140 GeV observed at low p_T (left) is a result of the subjet kinematics, where an angular separation of $\Delta R = 1.0$–1.5 leads to jet masses around this value. When changing the kinematics by relaxing the $p_\text{T,sub}$ requirement, the peak vanishes and a falling distribution is obtained. The width of this peak is reduced for intermediate ($400 < p_\text{T} < 600\,\text{GeV}$) transverse momenta (not shown) and a monotonically falling background distribution is obtained for values of $600 < p_\text{T} < 800\,\text{GeV}$ (right). Very similar distributions are obtained for values of $p_\text{T} > 800\,\text{GeV}$.

The distributions of the leading subjet's fractional transverse momentum $f_{p_\text{T}} = p_{\text{T},1}/p_\text{T}$ is shown in Fig. 2 (middle). Signal jets contain subjets with more evenly distributed transverse momenta, while for background jets the leading subjet carries a larger fractional p_T on average. The variable f_{p_T} shows good separation power between signal and background jets before a subjet multiplicity selection. After the

[3] For example, on a MacBook Pro with a 2.5 GHz Intel Core i5 processor and 16 GB 1600 MHz DDR3 Memory the runtime is about 25 ms per event for multiplicities of $\mathcal{O}(3000)$.

[4] While pile-up effects will worsen the overall performance of the algorithm, the change is not expected to be significant for moderate pile-up scenarios (up to 20–30 additional pile-up interactions).

[5] Final state partons are defined as partons which enter the hadronisation step.

[6] The VR algorithm alone affects the jet mass distribution similarly to a trimming [15] procedure for anti-k_t jets at high top quark p_T [18].

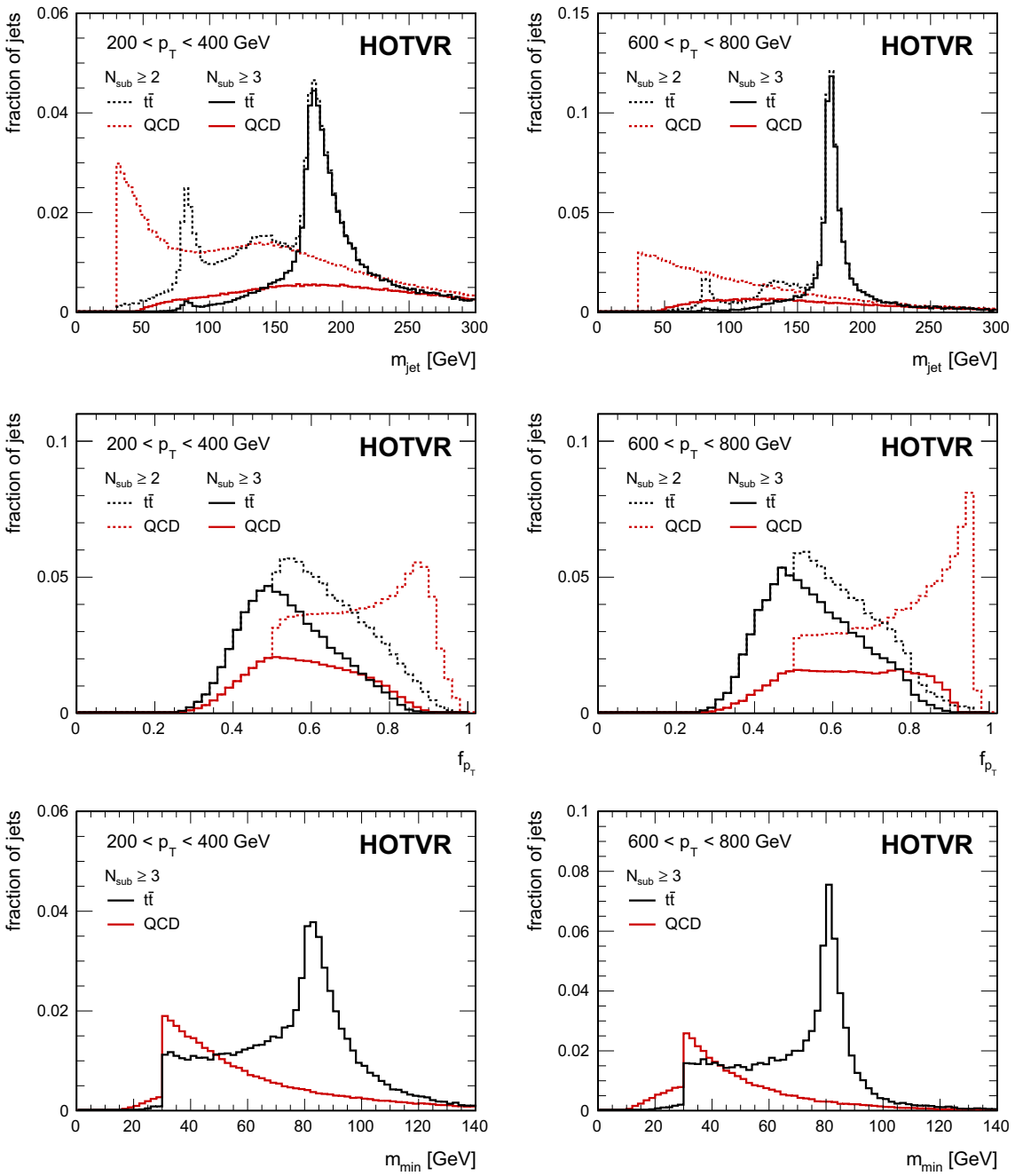

Fig. 2 Distribution of the jet mass (*top*), fractional leading subjet transverse momentum (*middle*) and minimum pairwise mass (*bottom*) for signal (*black*) and background (*red*) events as obtained with the HOTVR algorithm for two different ranges in parton jet p_{T}. The distributions are shown for subjet multiplicities $N_{\mathrm{sub}} \geq 2$ (*dashed lines*) and $N_{\mathrm{sub}} \geq 3$ (*solid lines*). Note that the minimum pairwise mass is only defined for $N_{\mathrm{sub}} \geq 3$. The distributions have been normalised to unit area for $N_{\mathrm{sub}} \geq 2$

requirement of $N_{\mathrm{sub}} \geq 3$, the separation power is reduced, but the variable is still useful, especially at high p_{T}.

For jets with $N_{\mathrm{sub}} \geq 3$, the distribution of the minimum pairwise mass m_{min} [26,27], defined as the minimum invariant mass of pairs of the three highest p_{T} subjets $m_{\mathrm{min}} = \min[m_{12}, m_{13}, m_{23}]$, is shown in Fig. 2 (bottom) for two regions of p_{T} of the parton jet. The distributions show a clear cut-off at the chosen value of the mass jump threshold ($\mu = 30\,\mathrm{GeV}$). Above this value the distribution is steeply falling for background jets, while $\mathrm{t\bar{t}}$ signal jets exhibit a pronounced peak around the value of the W boson mass, as expected for top quark jets. The tail below the mass jump

threshold is a result of light subjets combined with a heavier pseudojet, fulfilling the mass jump criterion in step (3) of the algorithm.

Besides an adequate reconstruction of masses, algorithms should also be able to reconstruct the kinematics of the initial heavy particle. In particular the size of the catchment area, which is responsible for the amount of additional radiation clustered into the jets, and the intensity of the grooming procedure are critical components for the performance in this area. For an evaluation of the kinematic object reconstruction by the HOTVR algorithm, we calculate the p_{T} ratio of the HOTVR jets and the matched parton jets containing a top quark. We find a mean value of the p_{T} ratio of 1.0 within small deviations of the order of 1%, independent of the parton jet p_{T}. The widths of the p_{T} ratio distributions are about 5%. This shows that the HOTVR algorithm is able to accurately reconstruct the kinematics of the heavy object with the parameter choice given above.

Selection cuts in top-tagging mode

For the discrimination of hadronically decaying top quarks from QCD multijets a selection based on simple cuts using commonly employed substructure variables has been implemented. The variables m_{jet} and m_{min} calculated from the HOTVR subjets are in principle sufficient for building a robust top tagger over a large region of p_{T}. However, cuts on additional variables have been added to obtain a selection that allows a fair comparison with other top-tagging algorithms using similar selections. To ensure only a limited impact of not-included experimental effects (e.g. broadening of distributions) the cut values have not been optimised rigorously. Nevertheless, they result in an improved discrimination between signal and background.[7] The following selection defines the standard working point of the HOTVR algorithm in top-tagging mode.

1. The leading subjet is required to have a fractional transverse momentum with respect to the jet, $f_{p_{\mathrm{T}}} = p_{\mathrm{T},1}/p_{\mathrm{T}} < 0.8$, which ensures that the jet's momentum is distributed among its subjets and not carried by only the leading subjet.
2. The number of subjets N_{sub} is required to be $N_{\mathrm{sub}} \geq 3$, which increases the probability of reconstructing fully merged top jets and rejects a fair amount of QCD jets.
3. The jet mass is required to fulfil $140 < m_{\mathrm{jet}} < 220\,\mathrm{GeV}$.
4. The minimum pairwise mass has to fulfil $m_{\mathrm{min}} > 50\,\mathrm{GeV}$.

These selection criteria lead to similar subjet kinematics as obtained by the CMS and HEPTopTagger algorithms with default parameters. This provides the basis for the comparison made in the following.

Performance comparison with ROC curves

The signal efficiency and misidentification rate are studied using single variable receiver operating characteristic (ROC) curves. The signal efficiency ε_S is defined as the fraction of tagged jets matched to parton jets containing the top quark, with respect to all top quarks decaying hadronically. The background efficiency (or misidentification rate) ε_B is calculated as the fraction of tagged jets matched to parton jets in a QCD multijet sample, with respect to the total number of parton jets. Both, ε_S and ε_B therefore combine identification and matching efficiencies. These definitions allow for a comparison of different tagging algorithms, in particular using different choices of the jet distance parameter R, since the reference p_{T} is defined by the parton jet matched to the tagged jet and does not depend on the specifics of the tagging algorithm under study.

In the following the performance of the HOTVR algorithm in top-tagging mode is compared with the performance of three top-tagging algorithms especially designed for dedicated regions of p_{T}: the CMS top-tagger targets the region of high p_{T}, the HEPTopTagger is designed for low p_{T} and its improved version with OptimalR has been developed to extend its usability to higher p_{T}. The free parameters of these taggers are listed in Table 3 together with a choice of working points [29,73,95]. The ROC curves are obtained by keeping the free parameters fixed at the values given and scanning only the N-subjettiness [21–23] ratio $\tau_{3/2} = \tau_3/\tau_2$ with $\beta = 1$. The choice of $\tau_{3/2}$ as scanning variable[8] ensures an unprejudiced comparison of the algorithms, which all rely on different reconstruction techniques and substructure variables, since this variable is not used in the definition of any of the taggers under study. Furthermore, $\tau_{3/2}$ has been shown to improve the performance of existing taggers (see for example Refs. [30,73]).

In Fig. 3 the ROC curves of the four top-tagging algorithms are shown for four different p_{T} regions, where p_{T} is defined by the parton jet matched to the tagged jet. The events were reweighted to obtain a flat p_{T} spectrum such that all events in the interval have the same weight. At low p_{T} ($200 < p_{\mathrm{T}} < 400\,\mathrm{GeV}$, top left) the CMS top-tagging algorithm has very small efficiency due to the choice of $R = 0.8$ which results in jets not large enough to cluster all particles from the top quark decay chain. The HOTVR

[7] A more sophisticated selection based on multivariate analysis techniques or more complex observables might provide further performance improvement [73] over this simple approach. However, the aim of the studies presented here is a comparison of the performance of the HOTVR algorithm with existing algorithms.

[8] The usual procedure for obtaining the ROC curves by scanning the free parameters of each algorithm could provide misleading results in this case, as it cannot be ensured that the usage of additional or different scanning variables for a given tagger would not improve its performance considerably.

Table 3 Settings of the top-tagging algorithms used. The parameter R is the distance parameter of the jet clustering. The definition of the parameters follows Ref. [95] for the CMS top tagger, Ref. [29] for the HEPTopTagger and Ref. [30] for the HEPTopTagger in OptimalR mode

CMS top tagger $R = 0.8$	HEPTopTagger $R = 1.5$		OptimalR $R = 0.5$–1.5
$\delta_p > 0.05$	$f_{\text{drop}} = 0.8$	$m_{23}/m_{123} > 0.35$	Same as HEPTopTagger
$A = 0.0004$	$m_{\text{cut}} = 30\,\text{GeV}$	$0.2 < \arctan \frac{m_{13}}{m_{12}} < 1.3$	$\Delta R = 0.1$
$N_{\text{sub}} \geq 3$	$R_{\text{filt}}^{\text{max}} = 0.3$	$f_W = 0.15$	$m_{\text{rec}}^{1.5} - m_{\text{rec}} > 0.2 m_{\text{rec}}^{1.5}$
$m_{\text{min}} > 50\,\text{GeV}$	$N_{\text{filt}} = 5$	$140 < m_{123} < 220\,\text{GeV}$	$\Delta R_{\text{opt}} < 0.5$
$140 < m_{\text{jet}} < 220\,\text{GeV}$	$p_{\text{T,sub}} > 30\,\text{GeV}$		

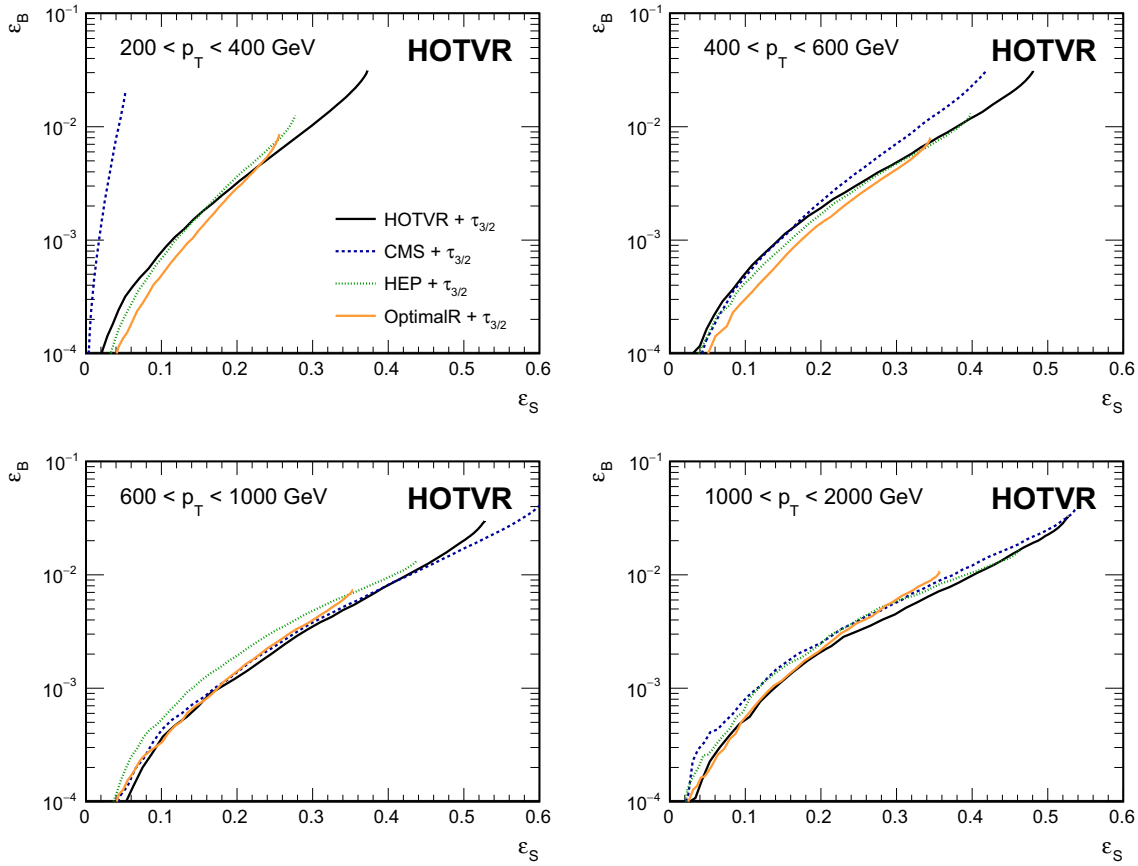

Fig. 3 Receiver operating characteristic (ROC) curves for different top-tagging algorithms obtained from a scan of the variable $\tau_{3/2}$ in four different p_{T} regions

algorithm is able to provide a comparable performance as the two HEPTopTagger algorithms which were optimised for this p_{T} region. For increasing values of p_{T} the CMS tagger becomes more efficient, with a similar performance as the OptimalR HEPTopTagger starting from $p_{\text{T}} > 600\,\text{GeV}$. In the p_{T} regions with $400 < p_{\text{T}} < 600\,\text{GeV}$ (top right) and $600 < p_{\text{T}} < 1000\,\text{GeV}$ (bottom left) the HOTVR algorithm shows a similar relation between ε_S and ε_B as the CMS and OptimalR HEPTopTagger, and is especially useful for high efficiencies. In the highest p_{T} region considered ($1000 < p_{\text{T}} < 2000\,\text{GeV}$, bottom right) the HOTVR algorithm features overall the best performance over all ε_S values, outperforming the CMS tagger, which was designed for the high p_{T} region.

In summary, the HOTVR algorithm shows a remarkably stable performance over a large range in p_{T} with similar or even better performance than algorithms especially designed for certain p_{T} regions. Detector reconstruction and resolution effects, which are not included in these studies, are expected to improve the performance of the HOTVR algorithm relative to the other algorithms studied [92].

5 Conclusion

A new algorithm for the reconstruction and identification of hadronically decaying heavy particles at the LHC has been introduced in this paper. The algorithm combines variable R jet clustering with a veto based on a mass jump criterion. It performs jet and subjet finding, and the rejection of soft radiation in one sequence. This combination results in a stable determination of jet substructure variables like the jet mass over a large range in p_{T} of the heavy object. In top-tagging mode the HOTVR algorithm provides an excellent ratio of signal to background efficiency at low top quark p_{T} as well as at high p_{T}, making the HOTVR algorithm useful in the regions of resolved and boosted decays at the same time.

While we focussed on top tagging in this work, the algorithm is also applicable for the tagging of W, Z, H or possible BSM resonances, where studies are ongoing. Because of its algorithmic simplicity combined with remarkable performance, this tagger could become a helpful ingredient for future boosted analyses at the LHC.

Acknowledgements We thank Michael Spannowsky for fruitful discussions during the development of the algorithm. We also thank Jesse Thaler for helpful suggestions on improvements of the document and for advise on speed improvements. This work is supported by the German Research Foundation (DFG) in the Collaborative Research Centre (SFB) 676 "Particles, Strings and the Early Universe" located in Hamburg.

Funded by SCOAP[3].

References

1. J.M. Butterworth, A.R. Davison, M. Rubin, G.P. Salam, Jet substructure as a new Higgs search channel at the LHC. Phys. Rev. Lett. **100**, 242001 (2008). arXiv:0802.2470
2. S.D. Ellis, C.K. Vermilion, J.R. Walsh, Techniques for improved heavy particle searches with jet substructure. Phys. Rev. D **80**, 051501 (2009). arXiv:0903.5081
3. G.P. Salam, Towards Jetography. Eur. Phys. J. C **67**, 637 (2010). arXiv:0906.1833
4. A. Abdesselam et al., Boosted objects: a probe of beyond the standard model physics. Eur. Phys. J. C **71**, 1661 (2011). arXiv:1012.5412
5. A. Altheimer et al., Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks. J. Phys. G **39**, 063001 (2012). arXiv:1201.0008
6. A. Altheimer et al., Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012. Eur. Phys. J. C **74**, 2792 (2014). arXiv:1311.2708
7. D. Adams et al, Towards an understanding of the correlations in jet substructure. Eur. Phys. J. C **75**, 409 (2015). arXiv:1504.00679
8. J.A.D.E. Collaboration, W. Bartel et al., Experimental studies on multi-jet production in e^+e^- annihilation at PETRA energies. Z. Phys. C **33**, 23 (1986)
9. J.A.D.E. Collaboration, S. Bethke et al., Experimental investigation of the energy dependence of the strong coupling strength. Phys. Lett. B **213**, 235 (1988)
10. M.R. Buckley, T. Plehn, M. Takeuchi, Buckets of tops. JHEP **08**, 086 (2013). arXiv:1302.6238
11. M.R. Buckley, T. Plehn, T. Schell, M. Takeuchi, Buckets of Higgs and tops. JHEP **02**, 130 (2014). arXiv:1310.6034
12. I.W. Stewart, F.J. Tackmann, J. Thaler, C.K. Vermilion, T.F. Wilkason, XCone: N-jettiness as an exclusive cone jet algorithm. JHEP **11**, 072 (2015). arXiv:1508.01516
13. J. Thaler, T.F. Wilkason, Resolving boosted jets with XCone. JHEP **12**, 051 (2015). arXiv:1508.01518
14. S.D. Ellis, C.K. Vermilion, J.R. Walsh, Recombination algorithms and jet substructure: pruning as a tool for heavy particle searches. Phys. Rev. D **81**, 094023 (2010). arXiv:0912.0033
15. D. Krohn, J. Thaler, L.-T. Wang, Jet trimming. JHEP **1002**, 084 (2010). arXiv:0912.1342
16. A.J. Larkoski, S. Marzani, G. Soyez, J. Thaler, Soft drop. JHEP **1405**, 146 (2014). arXiv:1402.2657
17. D. Krohn, J. Thaler, L.-T. Wang, Jets with variable R. JHEP **06**, 059 (2009). arXiv:0903.0392
18. ATLAS Collaboration, Boosted object tagging with variable-R jets in the ATLAS detector, ATL-PHYS-PUB-2016-013 (2016)
19. A.J. Larkoski, F. Maltoni, M. Selvaggi, Tracking down hyperboosted top quarks. JHEP **06**, 032 (2015). arXiv:1503.03347
20. J.M. Butterworth, B.E. Cox, J.R. Forshaw, WW scattering at the CERN LHC. Phys. Rev. D **65**, 096014 (2002). arXiv:hep-ph/0201098
21. I.W. Stewart, F.J. Tackmann, W.J. Waalewijn, N-Jettiness: an inclusive event shape to veto jets. Phys. Rev. Lett. **105**, 092002 (2010). arXiv:1004.2489
22. J. Thaler, K. Van Tilburg, Identifying boosted objects with N-subjettiness. JHEP **03**, 015 (2011). arXiv:1011.2268
23. J. Thaler, K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness. JHEP **02**, 093 (2012). arXiv:1108.2701
24. A.J. Larkoski, G.P. Salam, J. Thaler, Energy correlation functions for jet substructure. JHEP **06**, 108 (2013). arXiv:1305.0007
25. S.D. Ellis, A. Hornig, T.S. Roy, D. Krohn, M.D. Schwartz, Qjets: a non-deterministic approach to tree-based jet substructure. Phys. Rev. Lett. **108**, 182003 (2012). arXiv:1201.1914
26. D.E. Kaplan, K. Rehermann, M.D. Schwartz, B. Tweedie, Top tagging: a method for identifying boosted hadronically decaying top quarks. Phys. Rev. Lett. **101**, 142001 (2008). arXiv:0806.0848
27. CMS Collaboration, A Cambridge-Aachen (C-A) based jet algorithm for boosted top-jet tagging, CMS Physics Analysis Summary, CMS-PAS-JME-09-001, 2009
28. T. Plehn, G.P. Salam, M. Spannowsky, Fat jets for a light Higgs. Phys. Rev. Lett. **104**, 111801 (2010). arXiv:0910.5472
29. T. Plehn, M. Spannowsky, M. Takeuchi, D. Zerwas, Stop reconstruction with tagged tops. JHEP **1010**, 078 (2010). arXiv:1006.2833
30. G. Kasieczka, T. Plehn, T. Schell, T. Strebler, G.P. Salam, Resonance searches with an updated top tagger. JHEP **06**, 203 (2015). arXiv:1503.05921
31. D.E. Soper, M. Spannowsky, Finding physics signals with shower deconstruction. Phys. Rev. D **84**, 074002 (2011). arXiv:1102.3480
32. D.E. Soper, M. Spannowsky, Finding top quarks with shower deconstruction. Phys. Rev. D **87**, 054012 (2013). arXiv:1211.3140
33. ATLAS Collaboration, G. Aad et al, Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment. JHEP **05**, 128 (2012). arXiv:1203.4606

34. ATLAS Collaboration, G. Aad et al, ATLAS measurements of the properties of jets for boosted particle searches. Phys. Rev. D **86**, 072006 (2012). arXiv:1206.5369
35. ATLAS Collaboration, G. Aad et al, Performance of jet substructure techniques for large-R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. JHEP **09**, 076 (2013). arXiv:1306.4945
36. C.M.S. Collaboration, S. Chatrchyan et al., Studies of jet mass in dijet and W/Z + jet events. JHEP **05**, 090 (2013). arXiv:1303.4811
37. CMS Collaboration, Boosted top jet tagging at CMS, CMS Physics Analysis Summary, CMS-PAS-JME-13-007 (2014)
38. C.M.S. Collaboration, V. Khachatryan et al., Identification techniques for highly boosted W bosons that decay into hadrons. JHEP **12**, 017 (2014). arXiv:1410.4227
39. ATLAS Collaboration, G. Aad et al, Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**, 154 (2016). arXiv:1510.05821
40. ATLAS Collaboration, G. Aad et al, Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. arXiv:1603.03127
41. C.M.S. Collaboration, S. Chatrchyan et al., Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV. Phys. Lett. B **723**, 280 (2013). arXiv:1212.1910
42. C.M.S. Collaboration, S. Chatrchyan et al., Search for exotic resonances decaying into WZ/ZZ in pp collisions at $\sqrt{s} = 7$ TeV. JHEP **02**, 036 (2013). arXiv:1211.5779
43. C.M.S. Collaboration, S. Chatrchyan et al., Search for anomalous $\mathrm{t\bar{t}}$ production in the highly-boosted all-hadronic final state. JHEP **1209**, 029 (2012). arXiv:1204.2488
44. ATLAS Collaboration, G. Aad et al, Search for resonances decaying into top-quark pairs using fully hadronic decays in pp collisions with ATLAS at $\sqrt{s} = 7$ TeV. JHEP **01**, 116 (2013). arXiv:1211.2202
45. C.M.S. Collaboration, S. Chatrchyan et al., Search for top-quark partners with charge 5/3 in the same-sign dilepton final state. Phys. Rev. Lett. **112**, 171801 (2014). arXiv:1312.2391
46. CMS Collaboration, S. Chatrchyan et al, Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s} = 8$ TeV. Phys. Rev. Lett. **111**, 211804 (2013). arXiv:1309.2030. (Erratum ibid. 112 (2014) 119903)
47. ATLAS Collaboration, G. Aad et al, Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV, Phys. Rev. D **88**, 012004 (2013). arXiv:1305.2756
48. C.M.S. Collaboration, V. Khachatryan et al., Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV. JHEP **08**, 174 (2014). arXiv:1405.3447
49. ATLAS Collaboration, G. Aad et al, Search for $W' \to tb \to qqbb$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Eur. Phys. J. **C75**, 165 (2015). arXiv:1408.0886
50. ATLAS Collaboration, G. Aad et al, Search for resonant diboson production in the $\ell\ell$qq final state in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector. Eur. Phys. J. C **75**, 69 (2015). arXiv:1409.6190
51. CMS Collaboration, V. Khachatryan et al, Searches for third-generation squark production in fully hadronic final states in proton–proton collisions at $\sqrt{s} = 8$ TeV. JHEP **06**, 116 (2015). arXiv:1503.08037
52. CMS Collaboration, V. Khachatryan et al, Search for resonant $t\bar{t}$ production in proton–proton collisions at $\sqrt{s} = 8$ TeV. Phys. Rev. D **93**, 012001 (2016). arXiv:1506.03062
53. CMS Collaboration, V. Khachatryan et al, Search for vector-like charge 2/3 T quarks in proton–proton collisions at sqrt(s)= 8 TeV. Phys. Rev. D **93**, 012003 (2016). arXiv:1509.04177
54. CMS Collaboration, V. Khachatryan et al, Search for pair-produced vector-like B quarks in proton–proton collisions at $\sqrt{s} = 8$ TeV. arXiv:1507.07129
55. CMS Collaboration, V. Khachatryan et al, Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure. JHEP **06**, 080 (2015). arXiv:1503.01952
56. CMS Collaboration, V. Khachatryan et al, Search for the production of an excited bottom quark decaying to tW in proton–proton collisions at $\sqrt{s} = 8$ TeV. JHEP **01**, 166 (2016). arXiv:1509.08141
57. CMS Collaboration, V. Khachatryan et al, Search for narrow high-mass resonances in proton–proton collisions at $\sqrt{s} = 8$ TeV decaying to a Z and a Higgs Boson. Phys. Lett. B **748**, 255 (2015). arXiv:1502.04994
58. CMS Collaboration, V. Khachatryan et al, Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton–proton collisions at $\sqrt{s} = 8$ TeV. JHEP **02**, 145 (2016). arXiv:1506.01443
59. ATLAS Collaboration, G. Aad et al, Search for high-mass diboson resonances with boson-tagged jets in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **12**, 055 (2015). arXiv:1506.00962
60. ATLAS Collaboration, G. Aad et al, Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Eur. Phys. J. **C75**, 209 (2015). arXiv:1503.04677. (Erratum: Eur. Phys. J.C75,370(2015))
61. ATLAS Collaboration, G. Aad et al, A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **08**, 148 (2015). arXiv:1505.07018
62. ATLAS Collaboration, G. Aad et al, Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector. Eur. Phys. J. **C75**, 412 (2015). arXiv:1506.00285
63. ATLAS Collaboration, G. Aad et al, Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP **02**, 110 (2016). arXiv:1510.02664
64. ATLAS Collaboration, G. Aad et al, Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Phys. Rev. D **93**, 072007 (2016). arXiv:1510.06218
65. ATLAS Collaboration, G. Aad et al, Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector. Phys. Lett. B **755**, 285 (2016). arXiv:1512.05099
66. CMS Collaboration, V. Khachatryan et al, Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks. Eur. Phys. J. C (2016). arXiv:1602.08762 (Submitted to)
67. CMS Collaboration, V. Khachatryan et al, Search for massive WH resonances decaying into the $\ell\nu\mathrm{b\bar{b}}$ final state at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**, 237 (2016). arXiv:1601.06431
68. ATLAS Collaboration, G. Aad et al, Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector. arXiv:1602.05606
69. ATLAS Collaboration, G. Aad et al, Search for single production of a vector-like quark via a heavy gluon in the 4b final state with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV. arXiv:1602.06034
70. ATLAS Collaboration, G. Aad et al, Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s} = 8$ TeV proton–proton collisions using the ATLAS detector. arXiv:1510.03818
71. CMS Collaboration, V. Khachatryan et al, Measurement of the integrated and differential $t\bar{t}$ production cross sections for high-p_T top quarks in pp collisions at $\sqrt{s} = 8$ TeV. arXiv:1605.00116

72. CMS Collaboration, V tagging observables and correlations, CMS Physics Analysis Summary, CMS-PAS-JME-14-002 (2014)
73. CMS Collaboration, Top tagging with new approaches, CMS Physics Analysis Summary, CMS-PAS-JME-15-002 (2016)
74. ATLAS Collaboration, Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS, ATLAS Physics Note, ATL-PHYS-PUB-2015-033 (2015)
75. ATLAS Collaboration, Expected performance of boosted Higgs ($\rightarrow b\bar{b}$) Boson identification with the ATLAS detector at $\sqrt{s} =$ 13 TeV, ATLAS Physics Note, ATL-PHYS-PUB-2015-035 (2015)
76. ATLAS Collaboration, Boosted hadronic top identification at ATLAS for early 13 TeV data, ATLAS Physics Note, ATL-PHYS-PUB-2015-053 (2015)
77. M. Gouzevitch et al., Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production. JHEP **07**, 148 (2013). arXiv:1303.6636
78. M. Stoll, Vetoed jet clustering: the mass-jump algorithm. JHEP **04**, 111 (2015). arXiv:1410.4637
79. K. Hamaguchi, S.P. Liew, M. Stoll, How to decontaminate overlapping fat jets. Phys. Rev. D **92**, 015012 (2015). arXiv:1505.02930
80. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). arXiv:0802.1189
81. Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, JHEP **08**, 001 (1997). arXiv:hep-ph/9707323
82. M. Wobisch, 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). arXiv:hep-ph/9907280
83. S. Catani, Y.L. Dokshitzer, M.H. Seymour, B.R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. Nucl. Phys. B **406**, 187 (1993)
84. S.D. Ellis, D.E. Soper, Successive combination jet algorithm for hadron collisions. Phys. Rev. D **48**, 3160 (1993). arXiv:hep-ph/9305266
85. M. Cacciari, G.P. Salam, Dispelling the N^3 myth for the k_t jet-finder. Phys. Lett. B **641**, 57 (2006). arXiv:hep-ph/0512210
86. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097
87. Fastjet contrib. http://fastjet.hepforge.org/contrib. Accessed 11 Oct 2016
88. T. Sjostrand, S. Mrenna, P.Z. Skands, PYTHIA 6.4 Physics and Manual. JHEP **05**, 026 (2006). arXiv:hep-ph/0603175
89. T. Sjöstrand, S. Mrenna, P.Z. Skands, A brief introduction to PYTHIA 8.1. Comput. Phys. Commun. **178**, 852 (2008). arXiv:0710.3820
90. T. Sjöstrand et al., An introduction to pythia 8.2. Comput. Phys. Commun. **191**, 159 (2015). arXiv:1410.3012
91. M. Cacciari, G.P. Salam, G. Soyez, The catchment area of jets. JHEP **04**, 005 (2008). arXiv:0802.1188
92. T. Lapsien, Studies of top tagging identification methods and development of a new heavy object tagger. PhD thesis, Universität Hamburg (2016)
93. P. Skands, S. Carrazza, J. Rojo, Tuning PYTHIA 8.1: the Monash 2013 Tune. Eur. Phys. J. C **74**, 3024 (2014). arXiv:1404.5630
94. R.D. Ball et al., Parton distributions with LHC data. Nucl. Phys. B **867**, 244 (2013). arXiv:1207.1303
95. CMS Collaboration, Study of jet substructure in pp collisions at 7 TeV in CMS, CMS Physics Analysis Summary, CMS-PAS-JME-10-013 (2011)