JHEP09(2014)035

Hadronic b' search at the LHC with top and W taggers

Shuo Yang,[a] **Ji Jiang,**[b] **Qi-Shu Yan**[c,d,1] **and Xiaoran Zhao**[c]

[a] *Physics Department, Dalian University,*
Dalian, 116622, P.R. China
[b] *Physics Department, Zhengzhou University,*
Zhengzhou, 450001, P.R China
[c] *College of Physics Sciences, University of Chinese Academy of Sciences,*
Beijing 100039, P.R. China
[d] *Center for High-Energy Physics, Peking University,*
Beijing, 100871, P.R. China

E-mail: yangshuo@dlu.edu.cn, xiaotianer1986@sina.cn,
yanqishu@ucas.ac.cn, zxrlha@gmail.com

ABSTRACT: We study the sensitivity of a down type quark b' via process $pp \to b'\bar{b}' \to tW^-\bar{t}W^+$ using jet substructure methods at the LHC with the collision energy $\sqrt{s} = 14\,\mathrm{TeV}$. We consider the case that the b' is heavy (say from $800\,\mathrm{GeV}$ to $1500\,\mathrm{GeV}$) and concentrate on the feasibility of the full hadronic mode. Both top tagger (the HEP top tagger) and W tagger (the CMS W-tagging) are used to reconstruct all objects in the final states. In order to suppress huge SM background events and take into account various cases with different number of boosted objects, we propose a comprehensive reconstruction procedure so as to extract the most crucial observables of the signal events. When b' mass is $1\,\mathrm{TeV}$, it is found that with a $200\,\mathrm{fb}^{-1}$ dataset, the LHC may be able to detect the b' with a significance up to 10 or better. With a 3000 fb^{-1} dataset, the LHC may be able to probe the b' with a mass around up to $2\,\mathrm{TeV}$, only by using the hadronic mode.

KEYWORDS: Jets, Monte Carlo Simulations

---

[1]Corresponding author.

Contents

1   Introduction

The LHC collaborations have discovered a Higgs boson [1, 2], it is quite natural to ask what will be the next discovery that could be expected for future LHC runs. Extra quarks are one of the possible signals for the new physics, which are supposed to offer solutions to the fundamental issues of the standard model on electroweak symmetry breaking and mass generation of fundamental particles [3]. The heavy bottom like quarks have been predicted in Top-Coloron model [4] and top flavor seesaw [5], non-minimal supersymmetric extentions [6–9] and extra dimension models with warped space [10–13], etc.

     The extra quarks are good targets for future LHC runs due to their strong interaction with the particles of the SM, especially with gluons. The LHC can be called as a gluon-gluon machine due to the large gluon fluxes in highly accelerated protons. When kinematically accessible, these extra quarks can be copiously produced at the LHC. There are quite a few phenomenological studies for the feasibility of heavy quarks at the LHC, which can be found in literatures [14–18]. Meanwhile, signature of extra quarks are one of focus for experimental searches [19–25].

     In this paper, we focus on the search of vector-like b'.[1] Vector-like fermions do not contribute to "oblique parameters" in the leading order, and thus these parameters do not constrain their masses. However, the mixing angles between the vector-like fermions and the SM three generations fermions are required to be small because there is no GIM mechanism to suppress the FCNC related to these vector-like fermions. Currently, the most stringent limit on b' comes from CMS search [25]. Focusing on the strong pair-production

---

[1]The naive fourth generation chiral quarks have mostly been ruled out by the Higgs data [26, 27].

mechanism, CMS has set lower limits between 582 and 732 GeV on the vector-like b' quark mass for various decay branching ratios at 95% confidence level [25]. If the b' exclusively decays into a top quark and a W boson, considering the same sign lepton final state, the b' with mass below 732 GeV has been excluded at the 95% confidence level.

It is well-known that the LHC is a top quark factory and the discovery and precision measurements of $pp \rightarrow t\bar{t}$ in all the decay modes have a great significance to test the prediction of the SM. Among all the decay modes, the signal of the semi-leptonic mode is relatively easier to pick out, which enjoys a relative larger branching fraction and smaller SM backgrounds. The dileptonic mode has the smallest branching fraction but enjoys an even cleaner backgrounds [28, 29]. The fully hadronic mode is relatively difficult due to the large QCD and W + jets backgrounds and large uncertainties in determining the QCD activities. Nevertheless, measuring the fully hadronic mode is an important indispensable test of the SM prediction. The success of the measurement of fully hadronic mode at the Tevatron [30, 31] and the LHC [32] demonstrates that these detectors of hadronic machines are capable to detect the final state with a large multiplicity of jets and new physics with full hadronic final states [33, 34]. Recently, phenomenological studies using the fully hadronic mode to probe the $t\bar{t}H$ signal [35, 36], charged Higgs $tH^{\pm}$ signal [37] and top partner signal [38] have been done.

The early analysis on the search sensitivity of b' can be found in [14], where a simple W-jet mass method was used. By using the semileptonic and dileptonic modes, the authors found that 1 TeV heavy quark can be reachable with 100 fb^{-1} dataset. Another interesting work can be found in [15], where mainly using leptonic modes, Bob Holdom observed that it is difficult to use one cone to capture both boosted W boson and Top quark from b' decay. For the semileptonic mode studied in [16, 17], it is observed that once all physics objects (say two top quarks, two W bosons, two b' quarks) can be reconstructed it is possible to extract the most crucial variables (like the mass bump of b') to suppress the SM background to a controllable level. A thorough study for top-partner can be found in [18], where leptonic and b jet modes have been comprehensively analyzed. But the full hadronic mode has been left undone.

To our understanding, the study of hadronic mode of t' and b' has been untouched due to two major difficulties: 1) the full hadronic mode of the signal has high multiplicity of jets, and the combinatorics to find the characteristic parameters, like the masses of b' and t' are challenging; 2) Without characteristic variables for signal, it is difficult to distinguish signal from the SM multiple jet final states, say $t\bar{t}$ + jets. In the study for charged Higgs boson [37], two of our authors have noticed that the top tagger indeed can help to capture signal while maintaining suppression to the SM background even in full hadronic mode. It is observed that when the featured kinematic variables of signal are reconstructed, by using the multivariable analysis techniques, like the boost decision tree [39–41] and neural network analysis, it is possible to pick out sufficient signal events when luminosity is large enough (say 100 fb^{-1}). Taking into account the recent quick development in tagging the boost objects [42, 43], it is well-motivated to explore the hadronic mode by adopting the recently developed hadronic top quark taggers and W boson taggers.

When an extra quark is heavy, massive objects in its decay final states like top quarks,

W/Z/Higgs bosons, can be highly boosted. Using the jet substructure techniques [42, 43], it offers a promising method to pick out possible signals while keep good suppression to the SM multiple jet final states. Based on Monte Carlo methods, these jet substructure techniques have been demonstrated to work quite well in searching for top partners and bottom partners [44–50]. Recently, theoretical understanding on jet physics from QCD side has brought new insights to the jet substructure and related phenomenological analysis methods, such as jet functions [51], grooming [52], quark/gluon separation [53, 54] etc. For a comprehensive review on the jet substructure based on first principle QCD calculation and monte carlo tools, we refer to ref. [43].

In this work we use the hadronic top quark taggers and W boson taggers to explore the sensitivity of the LHC to b'. We consider the process for $pp \to b'\bar{b}'$ at LHC by assuming that b' decays to top quark and W boson 100% and study the heavy quark b' in the mass range $0.8\,\mathrm{TeV} < m_{b'} < 1.5\,\mathrm{TeV}$. We propose a reconstruction procedure and demonstrate how to reconstruct all physics objects in the signal events. We further use multivariable analysis methods to optimize cuts and explore the sensitivity of the LHC to b'.

This paper is organized as follows. In section II, we briefly review the boosted massive object taggers which will be used in this work. In section III, we present our phenomenological analysis. Finally, we make a discussion and give our conclusions in section IV.

2 Brief introduction to top taggers and W taggers

At the LHC, there are large samples of W and Z bosons, Higgs, and top quarks with a transverse momentum P_t that considerably exceeds their rest mass. In this kinematic regime, conventional reconstruction algorithms that rely on a one-to-one jet-to-parton assignment are often inappropriate, in particular for hadronic decays of such boosted objects. The technique of jet substructure has been developed to tag the boosted electroweak massive particles with hadronic decay [42, 43, 55]. Roughly speaking, these tagging algorithms work in two steps: firstly cluster jets with a much larger radius parameter to capture the energy of the complete hadronic decay in a single jet; secondly use delicate discriminating variables to anatomize the internal structure of these fat jets in order to separate boosted objects from the large QCD background. Below we summarize several most common used algorithms for boosted top and boosted W on the market.

2.1 Top-taggers

Top quarks play an important role in understanding electroweak symmetry breaking and searching for new physics. Unlike the case at Tevatron where most of top quarks are produced near the threshold, at the LHC many boosted top quarks can be produced. So, the jet substructure technique used to identify the boosted top from its hadronic decay has been developed in recent years. For a recent review on top taggers, we refer to ref. [56]. Here, we give a brief review on John Hopkins top-tagger [57] and the HEPTopTagger (Heidelberg-Eugene-Paris) [45, 46].

A1. Johns Hopkins top-tagger. After the success of BDRS algorithm [55] in Higgs search, Johns Hopkins group extended the BDRS algorithm to top study and proposed a top tagging method for the highly boosted top in hadronic decay [57]. We will dub it as "JHTopTagger" for simplicity and use it in later analysis. As well-known, a boosted top from its hadronic decay looks like a fat jet with three hard cores. Similar to the BDRS jet substructure method for boosted Higgs, the JHTopTagger firstly uses a large cone to cluster the event in order to capture all the decay products and relevant radiations of a boosted top and then de-clusters the top jet to find three subjets inside the massive mother jet.

To resolve a fat jet into the relevant hard substructure from top decay, the following recursive procedure are applied in JHTopTagger [57]:

1. Four momenta of all particles from top decay are clustered into a massive jet with a large cone size R (CA algorithm are used in original paper),

2. Undo the last combination to get two objects j_1 and j_2. If the P_t ratio of the softer jet j_2 over the original jet j is too small, i.e., $P_{t_{j_2}}/P_{t_j} < \delta_p$, throw the softer j_2 and go on to decluster on the harder one.

3. The declustering step is repeated until two separated hard objects are found. If any criterion below are satisfied, the declustering is failed: 1) both objects are softer than δ_p (2) two objects are too close, $\Delta\eta + \Delta\phi < \delta_r$ (3) the original jet is considered irreducible.

4. Declustering repeatedly on these two subjets will result in 2,3, or 4 hard objects.

5. It is required that the total mass of these subjets (only 3 or 4 hard subjets are considered) should be near m_t and the mass of two subjets among those resolved subjets should be in the m_W window. Furthermore, W helicity angle θ_t should be consistent with a top decay due to the left handedness of the SM. Here, the helicity angle θ_t is defined in the rest frame of the reconstructed W and is equal to the angle between the reconstructed top's fly-in direction and the fly-out direction of one of the two jets of W decay products. Typically, the softer subjet in the lab frame are chosen to set the angle.

The parameters involved in the method can be optimized event by event [57]. In our study, for the LHC context, these parameters are fixed as below:

$$\delta_p = 0.19 \qquad \delta_r = 0.1 \qquad |\theta_t| < 0.65 \ . \tag{2.1}$$

A2. HEPTopTagger. Similar to the Johns Hopkins tagger, the HEPTopTagger [45, 46] (Heidelberg-Eugene-Paris) is developed to capture moderately boosted top and is firstly used to improve $t\bar{t}H$ searches. It begins with a large $R = 1.5$ to cluster a fat CA jet. Such a large R allows us to access top quarks down to a lower $P_t \sim 200\,\text{GeV}$ at the price of a large combinatorics of subjets and serious pile-up problem. The HEPTopTagger uncluster the fat jet using an iterative mass-drop criterion. At the meantime, it employs

a filtering [55] procedure to pick up three hard subjets as candidates of top daughter jets and then test them with top kinematics. The detailed delicate steps in HEPTopTagger to capture boosted tops are given:

1. The last cluster of the fat jet j is undone to get j_1 and j_2. And then the mass drop criterion $\min m_{j_i} < \delta_m \cdot m_j$ determines if we keep j_1 and j_2. A subjet with a large jet mass $m_{j_i} > 30$ GeV are further decomposed; otherwise, the subjet is put to the list of relevant substructure.

2. The algorithm further uses the filtering procedure to construct one three-subjet combination with a jet mass closest to m_t as the top candidates.

3. If the three invariant masses (m_{12}, m_{13}, m_{23}) for the P_t ordering subjets j_1, j_2, j_3 satisfy one of the following three criteria, accept them as a top candidate:

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

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

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

$$(2.2)$$

4. For consistency, require the combined P_t of the three subjets to be above 200 GeV.

Here, the mass drop parameter δ_m and the mass windows parameters $R_{\min}$ and $R_{\max}$ are taken as $\delta_m = 0.8$, $R_{\min} = 85\% \times m_W/m_t$ and $R_{\max} = 115\% \times m_W/m_t$. The soft cutoff $R_{\text{soft}} = 0.35$ is supposed to remove QCD and W+jets background events. The HEPTopTagger has an identification efficiency of roughly 40% for top quarks with $P_t > 400$ GeV [45, 46].

We have compared the performance of these two top taggers and noticed that the HEPTopTagger can have a relative better performance, which can be attributed to the following two reasons: 1) the HEPTopTagger can capture not only highly boosted top but also intermediate boosted top; 2) due to more variables are used, the tagger can maintain a remarkable rejection to the SM background events even for the highly boosted top. Therefore, in the following study, we will adopt the HEPTopTagger to tag boosted top quarks in the signal events.

2.2 W-taggers

There are large samples of highly boosted electroweak massive particles, such as W bosons at the LHC. The hadronic decay products of these massive particles will be collimated to form fat jets. These W-jets are different from QCD jets in two main aspects. Firstly, a W jet contains two hard subjets in similar energy and mass, originated from the two quarks in the W decay, while a QCD jet usually has only one hard subjet and asymmetric energy-flow distribution. On the other hand, a QCD jet is initiated from a color triplet or

octet, which is color-connected to the beam or the other side of the event. Whereas, the two subjets of a W-jet are from a color singlet and they tend to correlate to each other in color. Along these lines, many sophisticated W-tagging tools has been developed to pick out the highly boosted W bosons from backgrounds [42, 58–60].

B1. CMS W tagger. The LHC experiment has employed the W tagging algorithms to search for new physics [58, 59]. Here, we briefly introduce the W tagging algorithm used by CMS collaboration (CMSWTagger) [59], which mainly use pruning [61, 62] and mass drop [55] methods. The algorithm can be applied to a massive jet and is given as follows.

1. The pruning method: the clustering history for a fat and massive jet is checked at every step. For a merging step say $(i + j \to p)$, two conditions are examined:

$$z_{ij} \equiv \frac{\min(P_{ti}, P_{tj})}{P_{tp}} > z_{\text{cut}} \tag{2.3}$$

and

$$R_{ij} < D_{\text{cut}} = 2 \times \frac{m_J}{P_{tJ}} \tag{2.4}$$

If this $(i + j \to p)$ step does not satisfy these two conditions, i and j will not be merged and instead the softer of the two clusters is removed.

2. The mass drop method: the total mass of the above pruned jets are required to be in the W mass window $70\text{GeV} < m_{\text{jet}} < 100\text{GeV}$. Undoing the last clustering iteration of the pruned jet to get two subjets. The ratio of masses of the hardest subjet (m_1) and the total pruned jet mass is defined as the mass drop $\mu = \frac{m_1}{m_{\text{jet}}}$. To discriminate against QCD jets, the mass drop is required to satisfy $\mu < 0.4$.

B2. Multivariate analysis W tagger. Unlike the case in top tagger, there are only few orthogonal variables in the two body decay of a highly boosted W. In ref. [60], a jet substructure algorithm with multivariate analysis was proposed for distinguishing highly boosted hadronically decaying W from QCD jets. The algorithm, dubbed it as "TMVAW-Tagger", selects 25 most useful variables and combines them using the Boosted Decision Trees method. These variables include the masses and P_t's after jet grooming, planar flows, P_t R-cores, etc. The detailed steps of the TMVAWTagger are presented as below:

1. Begin with fat jets with large $R = 1.2$ and then use the filtering/mass drop [55] to identify W jet candidates.

2. Apply a filtering step to get the leading three filtered subjets. The jet mass of combination of these jets is required to be within the mass window (60,100) GeV.

3. After the mass window cut, the original unfiltered fat jets are treated by using the multivariate analysis to maximize the efficiency. The 25 variables used in the analysis are:

$$m_{\text{jet}}, \ c_{P_t}(0.2 - 1.1), \ \text{sens}_{\text{filt,trim,prun}}^{m,P_t}, P_f, P_f(0.4), \frac{P_t^{\text{sub1,sub2}}}{P_t}, \frac{m^{\text{sub1,sub2}}}{m}, \ \Delta R_{\text{sub}}, n_{\text{sub}}. \tag{2.5}$$

Here, c_{P_t}'s are P_t R-cores [60], from 0.2 to 1.1 by 0.1 and $\text{sens}_{\text{filt,trim,prun}}^{m,P_t}$ represent 6 grooming [55, 61–63] sensitivities. P_f and $P_f(0.4)$ are the planar flow parameters for the original jet and for the highest P_t subjet from reclustering with $R = 0.4$, respectively. ΔR_{sub} is the distance between the two leading subjets and n_{sub} is the total number of subjets after the filtering process.

By using the multivariate analysis, the TMVAWTagger can be quite robust to reject background. Nonetheless, to simplify the current study, we use the CMSWTagger and choose a smaller cone to capture the W bosons from the decay of b'.

3 Numerical results and analysis

At the LHC, b' pair is mainly produced by the gluon fusion process. The cross section of $pp \rightarrow b'\bar{b}'$ has been studied at reference [64] up to NLO + NLL. For the collision energy $\sqrt{s} = 14\,\text{TeV}$, when $m_{b'}$ is around 1 (2) TeV, the cross section can be 69 (0.3) fb. When we assume that b' decays 100% to top and W boson, its decay width is less than 10%, so the narrow width approximation is still held.

3.1 Features of signal

In order to determine the right parameters to tag both top quarks and W bosons, we analyze the distributions of cone sizes of top quarks and W bosons in the final states at the parton level. We call the W bosons directly from the b' decay as the isolated W bosons and label them as W^{iso}. In contrast, we call the W bosons from the top quark decays as non-isolated W bosons and label them as W^{non}.

The correlations between the transverse momenta and the largest angle separation between two jets of two types of W bosons are shown in figure 1. In order to estimate the value of cone parameter R to cluster two jets from W boson hadronic decay into one fat jets, we define R_W^{iso} as $R_W^{\text{iso}} = \max(R(P(W), P(j_1)), R(P(W), P(j_2)))$, where $P(j_1)$ and $P(j_2)$ label the momenta of two daughter jets from the W boson decay and $P(W)$ labels the momentum of the W boson. From the plots, it is obvious that these two types of W bosons can be distinguished from their transverse momenta and angle separations. The most probable value of angle separation between two jets from the isolated W bosons is around 0.3 and is smaller than that from the non-isolated W bosons, which is around 0.6. The most probable transverse momenta of isolated W boson is around $500\,\text{GeV}$. It is larger than that of the non-isolated W bosons, which is around $300\,\text{GeV}$. These kinematic features can be utilized to determine the correct combinations of jets.

We also evaluate the largest angle separation of three partons from the top quark decay, which is defined as $R_t^{\text{max}} = \max(R(P(t), P(j_1)), R(P(t), P(j_2)), R(P(t), P(j_3)))$. The most probable value is around 0.8, which tells us that in order to capture the boosted top quarks, we'd better use a cone size parameter around 1.1 or so. By comparing the middle and right plots in the lower row given in figure 1, we can conclude that, at most of time, the $R(t)$ is determined by the angle separation of non-isolated W boson, which can also be read out from the left plot in the upper row from the curve $R_t^{\text{max}} - R_W^{\text{non}}$.
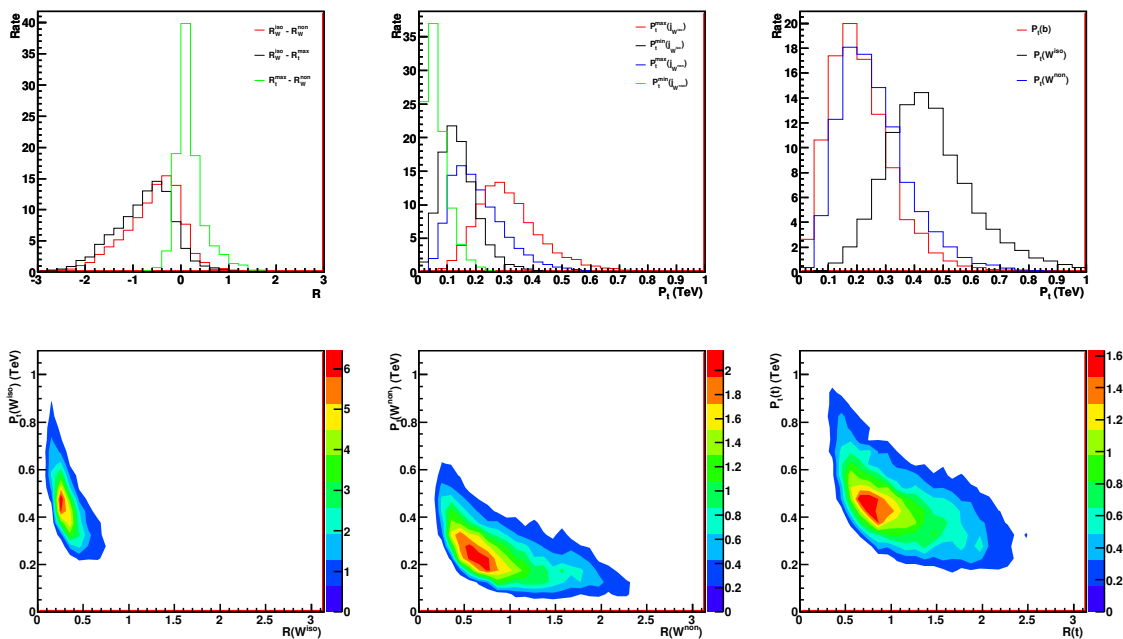
Figure 1. The momenta distributions and angle separations of jets from top quarks and W boson decay are demonstrated at parton level with $m_{b'} = 1\,\mathrm{TeV}$ as a show case.

We also examine the number of jets and one observable defined as $\sum \eta^2(j)$ with the change of the cone size parameters in the anti-kt algorithm, as demonstrated by figure 2. It is obvious that when the size of cone parameters for jet algorithm is changed, the number of jets can be changed, so as some kinematic observable, like $\sum \eta^2(j)$. There are also some observables, like the centrality, which are found to be insensitive to the change of cone parameter.

We also examine the jet mass distribution with the change of cone parameter R in the anti-Kt jet algorithm, as shown in figure 3. We would like to mark a few salient features from figure 3.

- When $R = 0.4$, it is found that more than 80% of the first leading jet can has a jet mass in the W-mass window (the window is defined as $|m(j) - m_W^{\mathrm{PDG}}| < 20$), while more than 50% of the second leading jet can have a jet mass in the W-mass window. Most of these massive jets are from the isolated W bosons, which is consistent with our analysis at the parton level shown in figure 1.

- When the cone parameter R is changed to 0.7, more than 40% of first leading massive jet has a jet mass in the top quark mass window (the window is defined as $|m(j) - m_t^{\mathrm{PDG}}| < 30$) and 30% of the first leading massive jet has a mass in the W-mass window. It is remarkable that more than 80% of the second and third leading massive jets are in the W-mass window. This indicates that the optimized cone parameters for W-jets should be around $R = 0.7$.
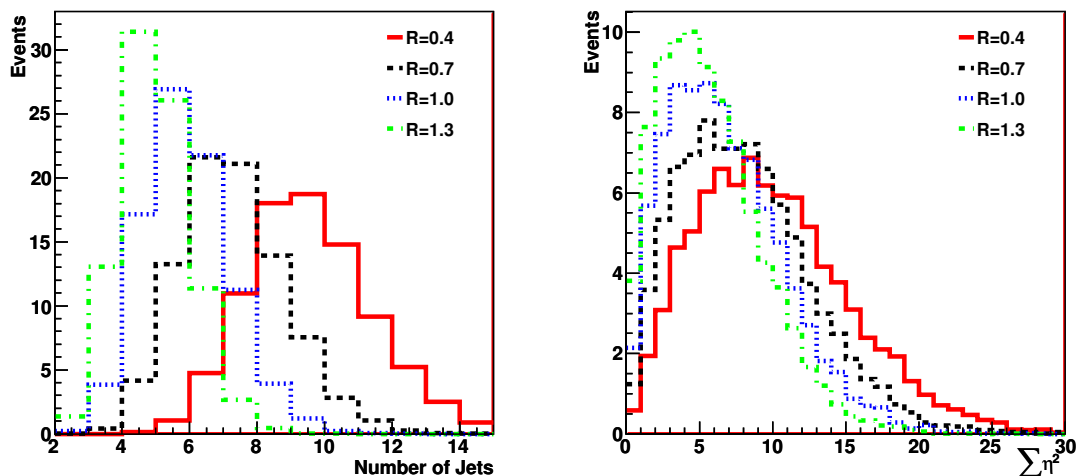
Figure 2. The number of jets and the observable $\sum \eta^2(j)$ changed with the cone parameter R in the anti-Kt jet algorithm (where detector effects are simulated by PGS) with $m_{b'} = 1\,\text{TeV}$ as our show case.

- When the cone parameter R is changed to 1.0, more than 50% of the first leading massive jet has a jet mass in the top quark mass window. There are around 25% of the second leading massive jet in the top quark mass window and 60% in the W-mass window. More than 85% of third leading massive jet is in the W-mass window. It is remarkable that there are around 40% the fourth massive jet in the W-mass window.

- When the cone parameter R is changed to 1.3, more massive jets can be in the top quark mass window. Compared with the plot for $R = 1.0$, we can read out that the optimized cone parameter R for a fat jet as a top quark should be around $1.0 < R < 1.4$ or so. Nevertheless, the optimized cone parameter should also take into account the behavior of background events. Another noticeable point is that the mass of the fifth jet indicates the mass dependence on the R.

Although some massive jets are outside the W-mass window or the top quark mass window, it is expected that the jet tagging techniques should help us to find their identities.

It is obvious that different R's can reveal parts of the full information of signal, a better way to separate signal and background is to utilize all available information with different R's when the computing time is allowed. When the computing resources are limited, we can use some tight preselection rules to choose the most relevant events in our analysis.

Obviously, when the mass of b' increases from $1\,\text{TeV}$ to $2\,\text{TeV}$, both W bosons and top quark become more energetic and their decay products can be more collimated. We observe that smaller cone size parameters can capture a considerable fraction of W bosons and top quarks, respectively, which is understandable from rule of thumb $R = 2\,m/P_t$.

It is remarkable that the origin of jet masses of top quarks and W bosons is different from that of QCD jet. The masses of those fat jets from top quark and W boson are from
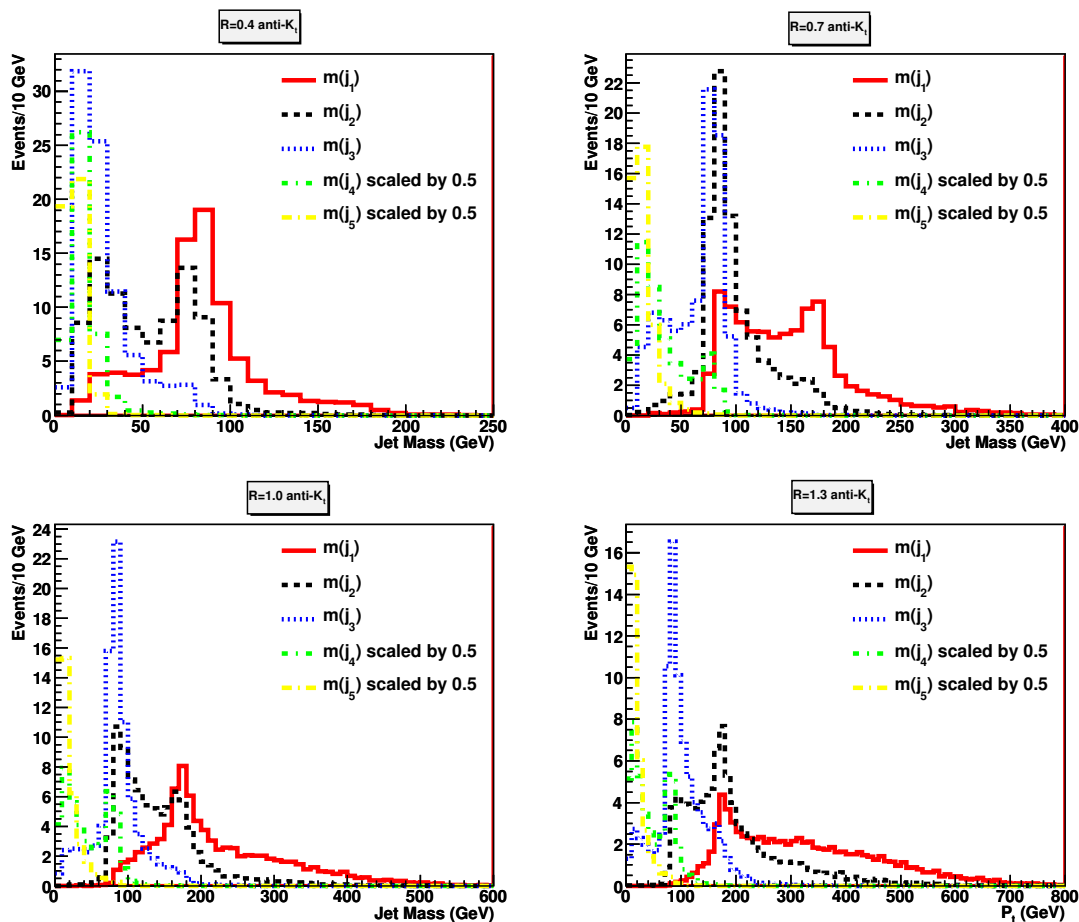
Figure 3. The jet mass distribution for the leading 5 massive jets in each event is shown, where jets are clustered by the anti-Kt jet algorithm (where detector effects are simulated by PGS) with $m_{b'} = 1\,\mathrm{TeV}$.

EW symmetry breaking, while the mass of QCD jet is from the collinear and infrared which can lead to wrong combination of pseudo-jets into a massive QCD jet.

In our Monte Carlo study, the signal events are generated by Madgraph/MadEvent [65] and background events by Alpgen [66]. We have used the MLM [67] matching to avoid the double counting issue. These events are fed to DECAY to generate full hadronic decay final states and pass to PYTHIA [68] to simulate showering, fragmentation/hadronization, initial state radiation, final state radiation, and multi-interaction as well. After that, fastjet [69, 70] and SpartyJet [71, 72] are used to perform jet clustering and massive object tagging analysis.

We would like to make a comment on the modeling of the SM background events. It is highly nontrivial to model the SM background events with 10 jets. For example, we noticed that the QCD jet sample is quite difficult to be generated by the MC tools on the market. Instead, by choosing the jet parameter $P_{t_{\min}} > 100\,\mathrm{GeV}$ in the Alpgen, we

	(0.5,1.3)	(0.6,1.2)	(0.6,1.3)	(0.6,1.4)	(0.7,1.2)	(0.7,1.3)	(0.7,1.4)	(0.8,1.4)
(0,0)	1	0	0	0	1	0	0	1
(0,1)	4	4	3	3	4	4	3	4
(1,0)	4	3	3	3	3	3	3	4
(0,2)	5	6	5	4	6	5	4	4
(0,3)	1	1	1	1	1	1	1	0
(1,1)	15	14	14	13	14	13	13	13
(1,2)	11	14	13	12	14	13	12	11
(1,3)	1	1	1	1	1	1	1	1
(2,0)	6	5	5	6	5	6	6	7
(2,1)	10	9	10	10	9	9	10	9
(2,2)	3	4	4	3	4	4	3	3

Table 1. The $(R_W\text{-}R_t)$-dependence of the fractions tagged hadronic top quark and tagged W boson in the process $pp \to b'\bar{b}' \to tW^-\bar{t}W^+$ are examined with $m_{b'} = 1\,\text{TeV}$ as a show case. At the head of each row, the first digit means the number of tagged top quark, and the second digit means the number of tagged W boson. For example, (1,2) means one tagged top quark and two tagged W bosons. For both top quark and W boson taggers, we adopt the CA jet algorithm. We use HEP top tagger for $R = 1.0 - 1.4$ to find tagged top quarks. We adopt the CMS W tagger to find tagged W bosons. Numbers in the table denote percentage. Numbers less than 1 are omitted.

generate exclusive datasets of $2j$, $3j$, $4j$, $5j$ events, and an inclusive $6j$ data sample. In our later analysis, we demand $n_j > 7$ and the P_t of the leading two jets momenta is larger than $200\,\text{GeV}$, so those extra jets can only be produced from the initial state radiation and finial state radiation. In this sense, for QCD jets data sample, our treatment can be regarded as leading order approximation. It is also true for case with the background events $W/Z+$ jets and diboson jets, where at most, Alpgen can allow us to generate $W/Z + 6$ jets at matrix element level. Nonetheless, we noticed that these two types of background can be efficiently suppressed by b taggings and the kinematics cuts introduced below. While for the $t\bar{t}$ dataset, we have used Alpgen to generate exclusive datasets $t\bar{t} + 0j$, $t\bar{t} + 1j, t\bar{t} + 2j$, $t\bar{t} + 3j$, and an inclusive dataset for $t\bar{t} + 4j$, and we merge these exclusive and inclusive datasets into an inclusive data for $t\bar{t}$ type background. For $t\bar{t}$ type background events, at most, we can marginally generate final states with 10 parton by the matrix elements. Keeping this fact in mind, the treatment to the SM background demonstrated in this work can only serve as a leading order approximation.

3.2 A proposed reconstruction procedure

To extract useful information of signals, we propose to cluster jets with three different sizes and the following reconstruction procedure to find all objects of an event:

1) For the small size jets, we use the anti-K$_t$ algorithm with jet parameter $R = 0.4$. We only consider high jet multiplicity events with $n_j \geq 9$ and $H_t > 1.5m_{b'}$. When there

are massive jets in the event, we demand the number of massive jets n_m (defined as $m(j_i) > 60$) and the number of non-massive jets should satisfy $2\,n_m + n_j \geq 10$.

2) For the mediate size jets, we use the CA algorithm with jet parameter $R = 0.6$. These mediate jets are supposed to find massive objects, especially the massive W bosons. Some of highly boosted top jets can be also found.

3) For the largest size jets, we use the CA algorithm with jet parameter $R = 1.3$, we find massive objects, especially the boosted top quarks. We label the number of tagged top quarks as n_t.

4) We identify non-isolated W bosons by using the massive jet found at step 2 and 3. We examine whether each of identified W bosons at step 2 is in the cone of the identified top quarks. If a tagged W boson is in the cone of a top jet, we label it as a non-isolated W boson. If not, it will be labelled as un-used and will be used to further determine the missing objects. The number of un-used tagged W bosons is denoted as n_W.

5) We identify isolated small jets with $R = 0.4$ which is neither in the cone of W bosons nor in the cone of top quarks. And we use them to reconstruct all missing objects, like top quark(s), W boson(s) and b's. To avoid the severe issue of wrong combinatorics, we throw away events when less than two objects are identified, i.e. $n_t + n_W < 2$.

There are several comments in order:

1) We use the minimum χ^2 approach to find the missing objects. For example, if in one event, we have tagged two top quarks with $n_t = 2$ and $n_W = 0$, then the rest of work is to reconstruct two W bosons by using the rest of isolated small jets. The χ^2 is defined as $\chi^2 = \frac{(m_{12}-m_W^{\text{PDG}})^2}{\sigma_W^2} + \frac{(m_{34}-m_W^{\text{PDG}})^2}{\sigma_W^2}$. If in one event, we have tagged two W bosons with $n_t = 0$ and $n_W = 2$, then the χ^2 is constructed as $\chi^2 = \frac{(m_{12}-m_W^{\text{PDG}})^2}{\sigma_W^2} + \frac{(m_{123}-m_t^{\text{PDG}})^2}{\sigma_t^2} + \frac{(m_{45}-m_W^{\text{PDG}})^2}{\sigma_W^2} + \frac{(m_{456}-m_t^{\text{PDG}})^2}{\sigma_t^2}$, where we have taken into account the possibility that any one of or both W bosons could come from top quark decays.

2) Once two top quarks and two W bosons have been reconstructed, we choose the value of $m^{\text{rec}} = m_{\text{min}}$ which can minimize the $\chi^2(m) = \frac{(m_{b'_1}-m)^2}{(\sigma_b')^2} + \frac{(m_{b'_2}-m)^2}{(\sigma_b')^2}$ as the reconstructed b' mass, where the mass of b'_i is the combination of a pair of top and W boson, where both two possible combinatorics have been taken into account.

From the results presented in table 1, we observe that if we require that $n_t = 2$ and $n_W = 2$, there are only $3-4\%$ of signal can be taken into account. While with $n_t + n_W \geq 2$ and the proposed reconstruction procedure, more than 50% of signals can be reconstructed successfully, while less than 1% of the dominant background $t\bar{t} + nj$ can pass through this reconstruction. These events can be correctly triggered at LHC collaborations due to both large H_t and large number of jets with the standard anti-K_t jet algorithm with the cone parameter $R = 0.4$. To further reduce reconstruction time, we choose $n_t + n_W \geq 2$ and $n_t \geq 1$ as our reconstruction conditions. It is found that by choosing this selection rule we can achieve similar results.

	$b'\bar{b}'$ [fb]	$t\bar{t}$ + jets [fb]	$t\bar{t}$ + W + jets [fb]
$\sigma \times$ branching fraction $\times$ btagging	6.0	1.67×10^5	55.6
$H_T > 1500\,\text{GeV}\,\&\,\mathcal{C} > 0.55\,\&\,n_j > 7$	4.2	2.3×10^3	2.5
Reconstruction $\&$ $m_{b'} > 650\,\text{GeV}$	2.2	147.2	0.2
MLP ($N > 0.9$)	1.5	3.7	0.02

Table 2. The cut efficiencies for signal and main background processes are shown here, where the signal is for the case $m_{b'} = 1\text{TeV}$. The cross section of $t\bar{t}$ is taken as 945 pb from [64], we assume all heavy particle decay in the hadronic mode. The optimized cut efficiencies for both MLP and BDT methods are provided for comparison. We cluster jets by using the standard anti-Kt algorithm with the cone parameter $R = 0.4$.

In table 1, fractions of tagged objects in the process $pp \rightarrow b'\bar{b}' \rightarrow tW^-\bar{t}W^+$ with different R are provided. This table provides important hints as how to capture our signal. There are a few comments in order. 1) The optimization in different cone sizes for R_W and R_t can affect the signal in 10% percentage level. Therefore, for different masses of b', it might be useful to optimize R_W and R_t. 2) It is crucial to capture top quark and W boson jets in our reconstruction procedure. 3) When R_W becomes too large (say larger than 0.8) , the fraction of signal events in the reconstruction procedure becomes fewer due to the failure of W-tagger.

3.3 Signal and background discrimination

In order to suppress the tremendous QCD background events and make hadronic mode doable, at the preselection level for further analysis, we demand that the scalar sum of transverse momentum of all final states must be larger than $\frac{3}{2}m_{b'}$ and the centrality of each event must be larger than 0.55. Furthermore, we require that at least two hadronic heavy objects (i.e. $n_t + n_W \geq 2$) must be identified and two b jets must be tagged. In this work, we assume that b tagging efficiency as 0.6 with a rejection factor 300. We further demand that there must be more than 7 jets with $P_t(j) > 20\,\text{GeV}$ and the transverse momentum of leading two jets should be larger than $200\,\text{GeV}$ in each event when jets are clustered with jet parameter $R = 0.4$. With these conditions, we observe that the background of QCD multiple jets are highly suppressed by the conditions of leading two jets $P_t(j_1) > 300\,\text{GeV}$ and $P_t(j_2) > 200\,\text{GeV}$, H_t, b taggings and jet numbers. Similarly, the background of $t\bar{t} + W +$ jets is also suppressed significantly by H_t and jet numbers. After these conditions for preselection, we find that the dominant background is $pp \rightarrow t\bar{t} +$ jets.

We have considered the contributions of diboson (WW, ZZ, etc.) backgrounds, $t\bar{t} + Z +$ jets, and have found that these types of reducible background events can be safely neglected after imposing preselection rules. Diboson background are also generated by Alpgen with the number of jets upto 3. Such type of background can also be heavily suppressed by the requirement of boosted objects and reconstruction criteria. Background events from $t\bar{t} + Z +$ jets with the number of jets upto 3 are very similar to $t\bar{t} + W +$ jets, which can at most reach to 1 percent of the signal after all cuts, therefore we neglect it here.
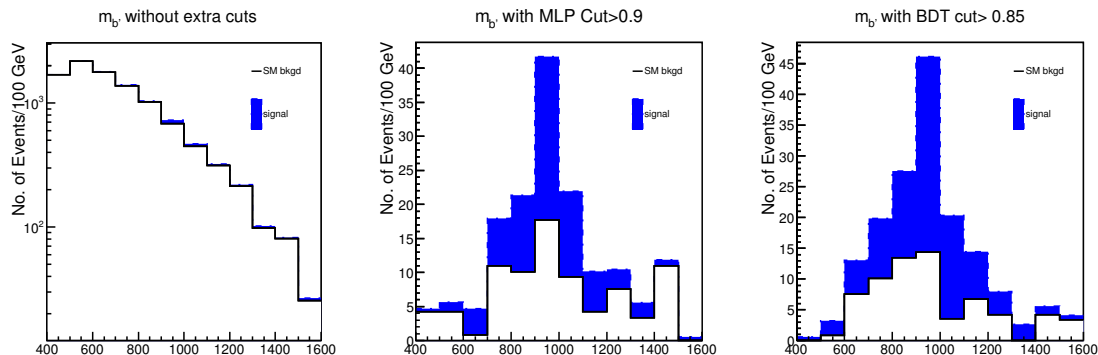
Figure 4. The distributions of reconstructed mass bump of b' before/after discrimination cuts are shown.

For the irreducible background $t\bar{t} + WW +$ jets with the number of jets upto 2 by using Madgraph5, we noticed that the cross section (3 fb for hadronic mode) of this type of background is close that of our signal when $m_{b'} = 1\,\text{TeV}$, but the H_t cut and the requirements in leading jets, the kinematics of reconstructed W bosons (two W bosons in our signal have very large P_t and are highly boosted), and the mass bump of b' (the background events occur near the threshold region) can help to remove such type of background down to a percent level of the signal after all cuts. Therefore, we neglect such type of background in table 2. For the same reasons, we also neglect $t\bar{t}t\bar{t}$ background with the number of jets upto 4 by using Alpgen, of which the cross section is 2 fb for fully hadronic mode.

In table 2, we illustrate how signal and background change with our preselection conditions. From it, it is observed that before reconstruction the dominant background is $pp \rightarrow t\bar{t} +$ jets. As demonstrated in left plot of figure 4, even after the reconstruction, it is still quite challenging to find the signal, the total background is around two order of magnitude larger than our signal. In order to achieve a better S/B and a better significance, obviously a dedicate signal background discrimination analysis is needed.

3.4 Multivariate analysis

Considering that there are 10 physics particles in our final state at parton level, the dimension of phase space of signal is 30 or so without taking into account the phase space of background events. As done in works [16, 17, 37], in this work we have adopted two Multivariate Analysis methods: the neural network (multilayer perceptron) and the boosted decision tree.

Below we roughly describe the crucial observables which can be used to distinguish signal and background. We can define the kinematic observables into two categories: observables without reconstruction and observables with reconstruction. The observables without reconstruction include observables which can be directly extracted from jets in the final state. For example, the transverse momentum and invariant mass of leading two jets can be obtained once we specify the jet clustering algorithms. Event shape observables, like the H_t, $\hat{s}$, centrality, and sphericity, can also be computed. As demonstrated in figure 5,
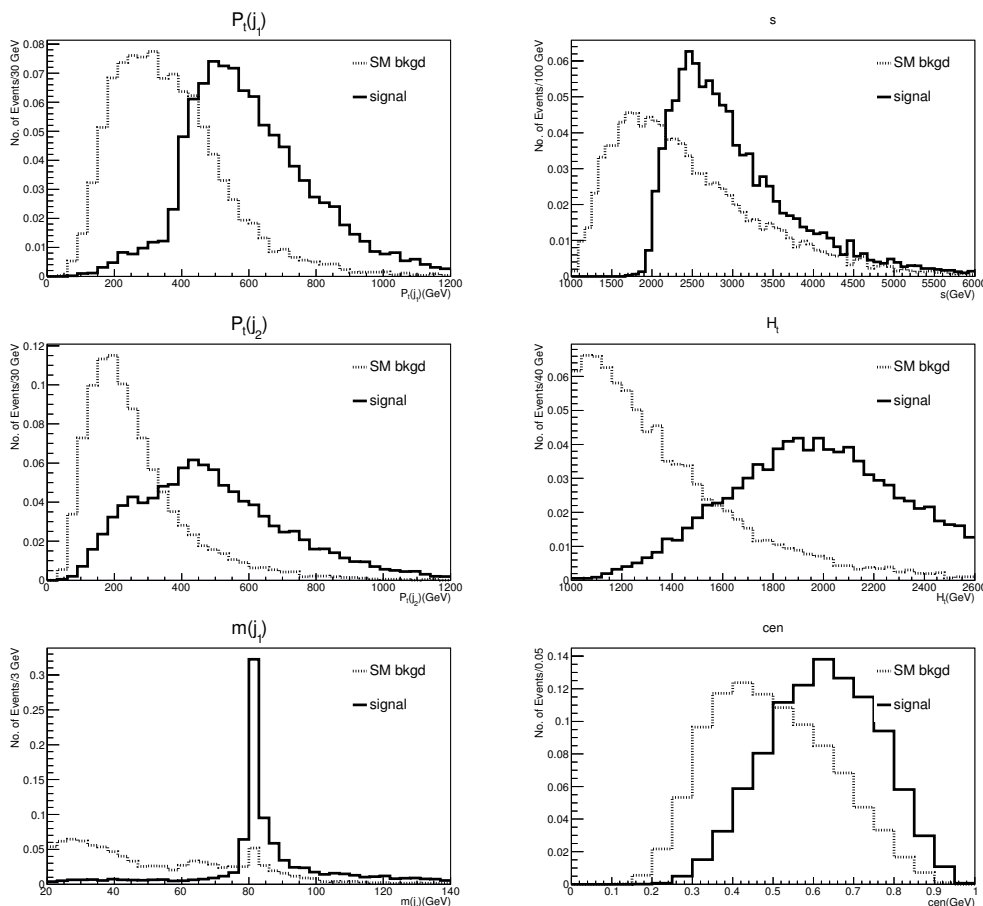
Figure 5. The observables showing the event shape and small jets for signal and background discrimination analysis are displayed, which can be obtained without using reconstruction procedure.

where the transverse momentum and invariant mass of leading two jets, the event shape variables, H_t, $\hat{s}$, and centrality are shown.

The second type of observables are those which can only be obtained after reconstruction. For example, the reconstructed masses, transverse momenta, and ηs of W bosons, top quarks and b' can only be obtained after we can identify all physics objects in term of our reconstruction procedure. In figure 6, we show the most useful and important observables which can help to discriminate signals and backgrounds, like the transverse momenta of reconstructed W bosons and top quarks, the transverse momentum of reconstructed b'.

For signal events, all these two types of observables are intrinsically correlated to the mass parameter of b'. In contrast, for the background events there is no such a correlation. By utlizing these observables and this correlation, we use the package TMVA to perform the training process and then apply the determined weights for each events which have not seen by the training process. We have applied both MLP neural network method and the boosted decision tree method [39–41]. The discriminant distributions for these
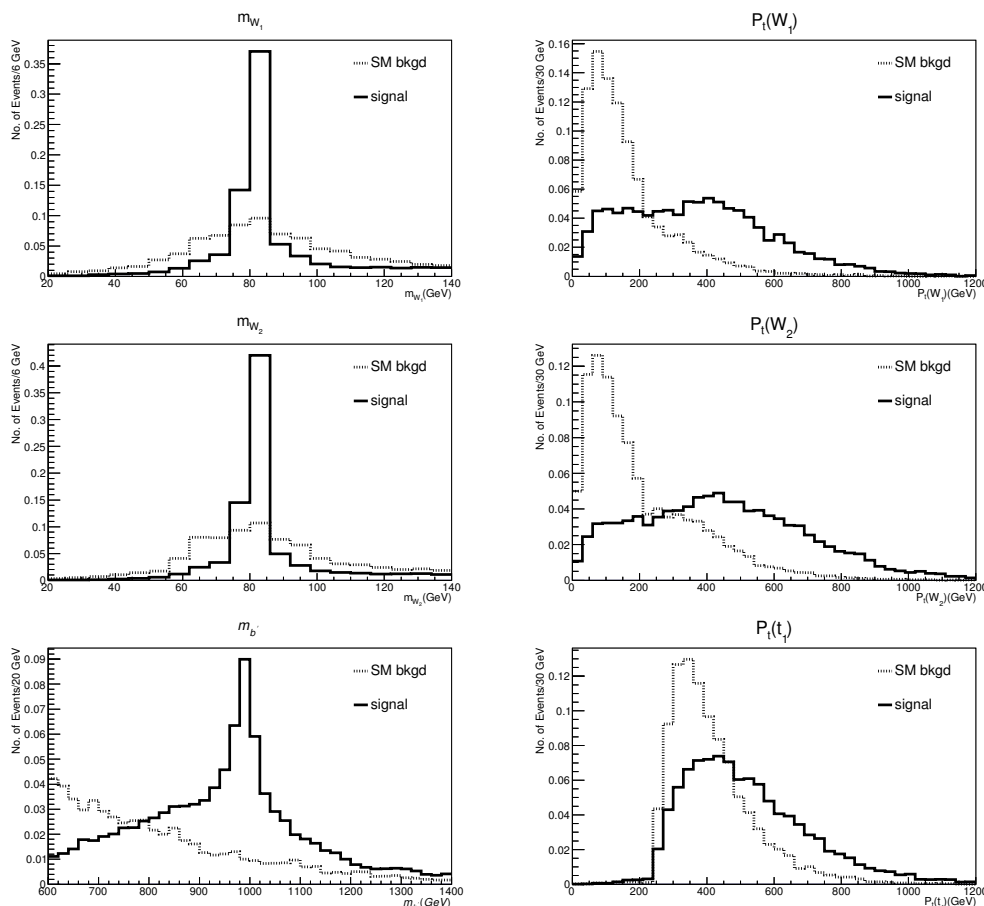
Figure 6. The observables for intrinsic theoretical parameters and phase space of signals are displayed, which can be obtained by using our proposed reconstruction procedure.

two MVA methods are provided in figure 7, which clearly demonstrate the discriminant analysis indeed works. We have also used the cut based method and observed that the MVA methods can optimize the signal and background discrimination better and improve the significance by a factor 100% or so, similar to the observation in our previous work [37] where the heavy charged Higgs boson search was studied. After using the MVA cuts, as demonstrated in figure 4, we can see the mass bump of reconstructed b' clearly standing out from the SM background.

3.5 Sensitivity of the LHC to b'

We use the analysis presented above to other values of $m_{b'}$ from $800\,\text{GeV}$ to $1500\,\text{GeV}$, where top quarks from b' decay can be either intermediately or highly boosted, and we arrive at the sensitivity given in table 3. We observe that when the b' is heavier its production rate becomes smaller, which leads to a smaller significance.

In table (4), we examine the upper bound on the cross section of $pp \to b'\bar{b}'$, where the significance $\sigma = 2.5$ is used to compute the exclusion bound.
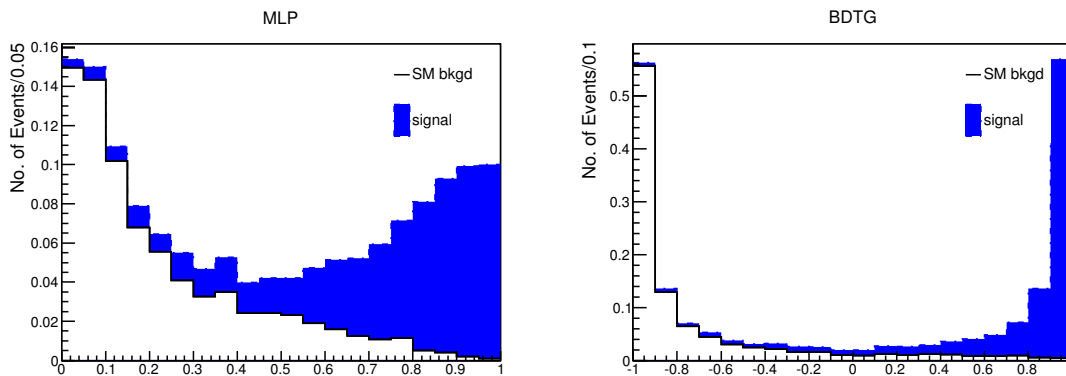
Figure 7. The signal and background discriminations in the MLP NN method and BDT method are demonstrated.

m'_b	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Signal	1316	708	356	209	108	58	38	14
Background	2560	1797	992	866	346	291	184	87
$\frac{S}{\sqrt{B}}$	26.0	16.7	11.3	7.1	5.8	3.4	2.8	1.5

Table 3. The significance achieved in our analysis is tabulated by varying the b' mass from $0.8\,\mathrm{TeV}$ to $1.5\,\mathrm{TeV}$. The optimized cone parameters for both W boson and top quark jets are provided in the first row. The number of signal and background are normalized to be $200\,\mathrm{fb}^{-1}$ with the collision energy $14\,\mathrm{TeV}$.

$m_{b'}$ (TeV)	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5
σ (fb)	288.2	137.2	69.0	36.2	19.7	11.0	6.3	3.7
$\frac{S}{\sqrt{B}}$ (with two b taggings and TMVA)	26.0	16.7	11.3	7.1	5.8	3.4	2.8	1.5
lower bound on $\sigma(fb)$	22.2	16.4	12.3	10.2	6.8	6.5	4.5	4.9

Table 4. The significance and sensitivity of LHC for b' production are shown, where we assume the total integrated luminosity as $200\,\mathrm{fb}^{-1}$. We use $\frac{S}{\sqrt{B}} = 2.5$ to define the exclusion upper bound on the cross section.

In figure 8, we present the sensitivity of the LHC to the hadronic b' mode with integrated luminosity 200 fb^{-1} and $3000\,\mathrm{fb}^{-1}$, respectively. From this plot, we can estimate that with $3000\,\mathrm{fb}^{-1}$ dataset just using the hadronic b' mode, we can either find or rule out a b' up to 1.8-2.0 TeV or so. Compared with the semileptonic modes analyzed in [14], it is expected that the hadronic mode demands more luminosity due to the large SM background events, similar to the case of the hadronic mode of $t\bar{t}$.

4 Discussions and conclusions

In this paper, we have studied the full hadronic mode for the process $pp \to b'b' \to tW^-\bar{t}W^+$ at the LHC 14 TeV collision. The main task is to reconstruct the b' from the large com-
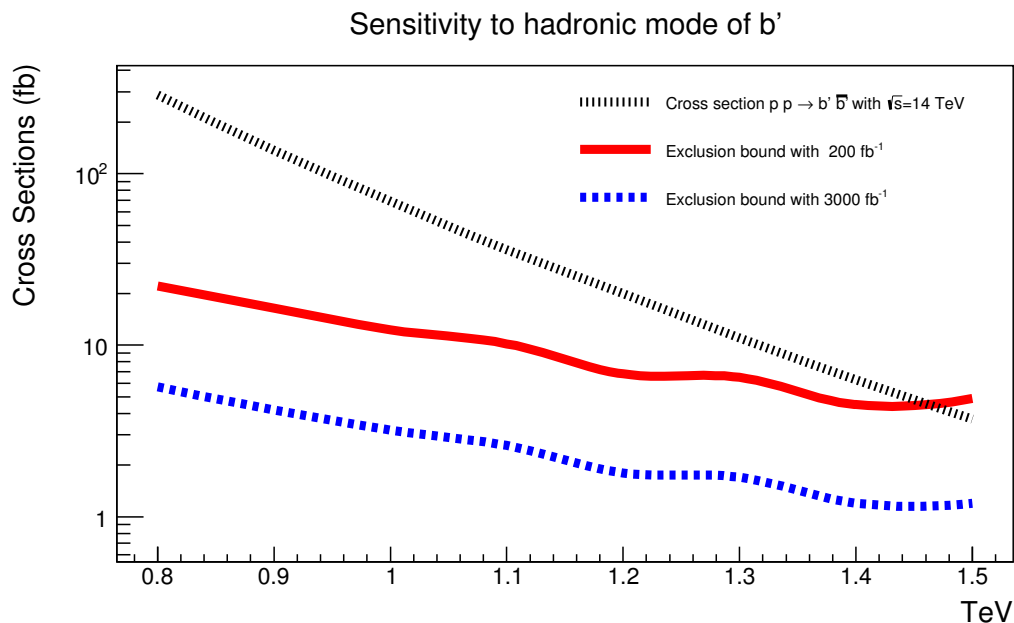
Figure 8. The sensitivity to the signal of hadronic b' mode at LHC 14 TeV is shown. The results for 3000 fb^{-1} dataset is obtained by directly scaling up the results of 200 fb^{-1}.

binatorics and to suppress the huge QCD and $t\bar{t}+$ jets background events, where we have found that b taggings are essential to suppress background events with large jet multiplicity from QCD. By using the top-tagger and W-tagger and some comprehensible cuts, we propose a full reconstruction procedure and demonstrate that if we could reconstruct the most important parameters of the signal, like the mass of b' and transverse momentum of W bosons, etc., the hadronic mode of b' could be feasible for the future LHC runs.

The current work can be directly extended for the higher energy collisions (say 100 TeV collisions). Obviously, for the signal with a fixed mass (say $m'_b = 10$ TeV), the signal will be enhanced due to the increase of collision energy and the corresponding large enhancement in the gluon fluxes. Nonetheless, more background processes may become dominant, for example, the multiple W boson final states, $t\bar{t} + W$, $t\bar{t} + WW$, $t\bar{t}t\bar{t}$, and tW. Some of these backgrounds could have significantly large enhancement in cross sections, say $t\bar{t}t\bar{t}$ and $t\bar{t}h$. Moreover, the collimated W bosons from EW showers [73, 74] in energetic b jets might fake the top quark taggers to some degree, which might be crucial for high energy collisions. Another challenging issue for 100 TeV collisions study is to model the multijet background events from the SM. We leave a detailed analysis for high energy collisions in our future works.

To project the sensitivity for b' at higher energy collisions, it might be useful to take into account leptonic modes. If so, a top tagger with leptonic modes should be useful [47]. When combining both leptonic and hadronic modes, we expect that higher exclusion bounds for b' can be achieved or a better significance can be obtained if a b' is there.

We have not included the pileup effects to the top and W taggers here, which is necessary for a more realistic analysis at either LHC future runs or future high energy

collisions. As demonstrated in the reference [75], the pileup effects might decrease the significance to a certain degree. Another interesting thing for high energy collisions is that when a top quark is very highly boosted (say $P_t > 5\,\mathrm{TeV}$), the current top tagger should be improved as demonstrated in [76] by taking into account additional information from tracker system, or it might be improved from hardwares, say by increasing the granularity of detectors from 0.1×0.1 to 0.01×0.01. Obviously, studying boosted physics objects and their taggings at high energy can help us in the detector designs for future high energy collisions.

Acknowledgments

References

[1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B 716** (2012) 1 [arXiv:1207.7214] [INSPIRE].

[2] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B 716** (2012) 30 [arXiv:1207.7235] [INSPIRE].

[3] A. De Simone, O. Matsedonskyi, R. Rattazzi and A. Wulzer, *A First Top Partner Hunter's Guide*, *JHEP* **04** (2013) 004 [arXiv:1211.5663] [INSPIRE].

[4] R.S. Chivukula, E.H. Simmons and N. Vignaroli, *A Flavorful Top-Coloron Model*, *Phys. Rev.* **D 87** (2013) 075002 [arXiv:1302.1069] [INSPIRE].

[5] X.-F. Wang, C. Du and H.-J. He, *LHC Higgs Signatures from Topflavor Seesaw Mechanism*, *Phys. Lett.* **B 723** (2013) 314 [arXiv:1304.2257] [INSPIRE].

[6] S.P. Martin, *Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry*, *Phys. Rev.* **D 81** (2010) 035004 [arXiv:0910.2732] [INSPIRE].

[7] S.P. Martin, *Raising the Higgs mass with Yukawa couplings for isotriplets in vector-like extensions of minimal supersymmetry*, *Phys. Rev.* **D 82** (2010) 055019 [arXiv:1006.4186] [INSPIRE].

[8] C. Liu, *Supersymmetry and Vector-like Extra Generation*, *Phys. Rev.* **D 80** (2009) 035004 [arXiv:0907.3011] [INSPIRE].

[9] C. Liu and J.-S. Lu, *Dark Matter and Gauge Coupling Unification in A Supersymmetry Model with Vector-like Matter*, *JHEP* **05** (2013) 040 [arXiv:1305.0070] [INSPIRE].

[10] K. Agashe, R. Contino, L. Da Rold and A. Pomarol, *A Custodial symmetry for Zbb̄*, *Phys. Lett.* **B 641** (2006) 62 [hep-ph/0605341] [INSPIRE].

[11] S. Casagrande, F. Goertz, U. Haisch, M. Neubert and T. Pfoh, *The Custodial Randall-Sundrum Model: From Precision Tests to Higgs Physics*, *JHEP* **09** (2010) 014 [arXiv:1005.4315] [INSPIRE].

[12] S. Gopalakrishna, T. Mandal, S. Mitra and R. Tibrewala, *LHC Signatures of a Vector-like b'*, *Phys. Rev.* **D 84** (2011) 055001 [arXiv:1107.4306] [INSPIRE].

[13] S. Gopalakrishna, T. Mandal, S. Mitra and G. Moreau, *LHC Signatures of Warped-space Vectorlike Quarks*, arXiv:1306.2656 [INSPIRE].

[14] W. Skiba and D. Tucker-Smith, *Using jet mass to discover vector quarks at the LHC*, *Phys. Rev.* **D 75** (2007) 115010 [hep-ph/0701247] [INSPIRE].

[15] B. Holdom, *The Heavy quark search at the LHC*, *JHEP* **08** (2007) 069 [arXiv:0705.1736] [INSPIRE].

[16] B. Holdom and Q.-S. Yan, *The B' Search at the LHC*, *Phys. Rev.* **D 84** (2011) 094012 [arXiv:1004.3031] [INSPIRE].

[17] B. Holdom and Q.-S. Yan, *Searches for the t' of a fourth family*, *Phys. Rev.* **D 83** (2011) 114031 [arXiv:1101.3844] [INSPIRE].

[18] J.A. Aguilar-Saavedra, *Identifying top partners at LHC*, *JHEP* **11** (2009) 030 [arXiv:0907.3155] [INSPIRE].

[19] ATLAS collaboration, *Search for down-type fourth generation quarks with the ATLAS detector in events with one lepton and hadronically decaying W bosons*, *Phys. Rev. Lett.* **109** (2012) 032001 [arXiv:1202.6540] [INSPIRE].

[20] ATLAS collaboration, *Search for pair production of a new quark that decays to a Z boson and a bottom quark with the ATLAS detector*, *Phys. Rev. Lett.* **109** (2012) 071801 [arXiv:1204.1265] [INSPIRE].

[21] CMS collaboration, *Search for heavy bottom-like quarks in 4.9 inverse femtobarns of pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **05** (2012) 123 [arXiv:1204.1088] [INSPIRE].

[22] CMS collaboration, S. Bhattacharya, *Search for Exotic Top Partners at $\sqrt{s} = 8$ TeV*, arXiv:1310.2299 [INSPIRE].

[23] CMS collaboration, *Search for top-quark partners with charge 5/3 in the same-sign dilepton final state*, *Phys. Rev. Lett.* **112** (2014) 171801 [arXiv:1312.2391] [INSPIRE].

[24] CMS collaboration, *Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* **B 729** (2014) 149 [arXiv:1311.7667] [INSPIRE].

[25] CMS collaboration, *Search for pair-produced vector-like quarks of charged -1/3 in lepton+jets final state in pp collisions at $\sqrt{s} =8$ TeV*, CMS-PAS, B2G-12-019.

[26] A. Djouadi and A. Lenz, *Sealing the fate of a fourth generation of fermions*, *Phys. Lett.* **B 715** (2012) 310 [arXiv:1204.1252] [INSPIRE].

[27] O. Eberhardt et al., *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*, *Phys. Rev. Lett.* **109** (2012) 241802 [arXiv:1209.1101] [INSPIRE].

[28] CDF collaboration, F. Abe et al., *Observation of top quark production in p̄p collisions*, *Phys. Rev. Lett.* **74** (1995) 2626 [hep-ex/9503002] [INSPIRE].

[29] D0 collaboration, S. Abachi et al., *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) 2632 [hep-ex/9503003] [INSPIRE].

[30] D0 collaboration, V.M. Abazov et al., *Measurement of the $p\bar{p} \to t\bar{t}$ production cross section at $\sqrt{s}$ = 1.96-TeV in the fully hadronic decay channel.*, *Phys. Rev.* **D 76** (2007) 072007 [hep-ex/0612040] [INSPIRE].

[31] CDF collaboration, T. Aaltonen et al., *Measurement of the $p\bar{p} \to t\bar{t}$ production cross- section and the top quark mass at $\sqrt{s}$ = 1.96-TeV in the all-hadronic decay mode*, *Phys. Rev.* **D 76** (2007) 072009 [arXiv:0706.3790] [INSPIRE].

[32] CMS collaboration, *Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s}$=7 TeV*, *Eur. Phys. J.* **C 74** (2014) 2758 [arXiv:1307.4617] [INSPIRE].

[33] CDF collaboration, T. Aaltonen et al., *Search for resonant production of $t\bar{t}$ decaying to jets in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV*, *Phys. Rev.* **D 84** (2011) 072003 [arXiv:1108.4755] [INSPIRE].

[34] CMS collaboration, *Search for Anomalous $t\bar{t}$ Production in the Highly-Boosted All-Hadronic Final State*, *JHEP* **09** (2012) 029 [arXiv:1204.2488] [INSPIRE].

[35] M.R. Buckley, T. Plehn and M. Takeuchi, *Buckets of Tops*, *JHEP* **08** (2013) 086 [arXiv:1302.6238] [INSPIRE].

[36] M.R. Buckley, T. Plehn, T. Schell and M. Takeuchi, *Buckets of Higgs and Tops*, *JHEP* **02** (2014) 130 [arXiv:1310.6034] [INSPIRE].

[37] S. Yang and Q.-S. Yan, *Searching for Heavy Charged Higgs Boson with Jet Substructure at the LHC*, *JHEP* **02** (2012) 074 [arXiv:1111.4530] [INSPIRE].

[38] M. Endo, K. Hamaguchi, K. Ishikawa and M. Stoll, *Reconstruction of Vector-like Top Partner from Fully Hadronic Final States*, arXiv:1405.2677 [INSPIRE].

[39] B.P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu et al., *Boosted decision trees, an alternative to artificial neural networks*, *Nucl. Instrum. Meth.* **A 543** (2005) 577 [physics/0408124] [INSPIRE].

[40] H.-J. Yang, B.P. Roe and J. Zhu, *Studies of boosted decision trees for MiniBooNE particle identification*, *Nucl. Instrum. Meth.* **A 555** (2005) 370 [physics/0508045] [INSPIRE].

[41] H.-J. Yang, B.P. Roe and J. Zhu, *Studies of stability and robustness for artificial neural networks and boosted decision trees*, *Nucl. Instrum. Meth.* **A 574** (2007) 342 [INSPIRE].

[42] 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].

[43] 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** (2014) 2792 [arXiv:1311.2708] [INSPIRE].

[44] G.D. Kribs, A. Martin and T.S. Roy, *Higgs boson discovery through top-partners decays using jet substructure*, *Phys. Rev.* **D 84** (2011) 095024 [arXiv:1012.2866] [INSPIRE].

[45] T. Plehn, G.P. Salam and M. Spannowsky, *Fat Jets for a Light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [arXiv:0910.5472] [INSPIRE].

[46] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops*, *JHEP* **10** (2010) 078 [arXiv:1006.2833] [INSPIRE].

[47] T. Plehn, M. Spannowsky and M. Takeuchi, *Boosted Semileptonic Tops in Stop Decays*, *JHEP* **05** (2011) 135 [arXiv:1102.0557] [InSPIRE].

[48] B. Bhattacherjee, S.K. Mandal and M. Nojiri, *Top Polarization and Stop Mixing from Boosted Jet Substructure*, *JHEP* **03** (2013) 105 [arXiv:1211.7261] [InSPIRE].

[49] A. Azatov, M. Salvarezza, M. Son and M. Spannowsky, *Boosting Top Partner Searches in Composite Higgs Models*, *Phys. Rev.* **D 89** (2014) 075001 [arXiv:1308.6601] [InSPIRE].

[50] A. Chakraborty, D.K. Ghosh, D. Ghosh and D. Sengupta, *Stop and sbottom search using dileptonic M_{T2} variable and boosted top technique at the LHC*, *JHEP* **10** (2013) 122 [arXiv:1303.5776] [InSPIRE].

[51] C.F. Berger, T. Kucs and G.F. Sterman, *Event shape/energy flow correlations*, *Phys. Rev.* **D 68** (2003) 014012 [hep-ph/0303051] [InSPIRE].

[52] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [arXiv:1307.0007] [InSPIRE].

[53] H.-n. Li, Z. Li and C.-P. Yuan, *QCD resummation for jet substructures*, *Phys. Rev. Lett.* **107** (2011) 152001 [arXiv:1107.4535] [InSPIRE].

[54] H.-n. Li, Z. Li and C.-P. Yuan, *QCD resummation for light-particle jets*, *Phys. Rev.* **D 87** (2013) 074025 [arXiv:1206.1344] [InSPIRE].

[55] J.M. Butterworth, A.R. Davison, M. Rubin and G.P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [arXiv:0802.2470] [InSPIRE].

[56] T. Plehn and M. Spannowsky, *Top Tagging*, *J. Phys.* **G 39** (2012) 083001 [arXiv:1112.4441] [InSPIRE].

[57] D.E. Kaplan, K. Rehermann, M.D. Schwartz and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, *Phys. Rev. Lett.* **101** (2008) 142001 [arXiv:0806.0848] [InSPIRE].

[58] CMS collaboration, *Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV*, *Phys. Lett.* **B 723** (2013) 280 [arXiv:1212.1910] [InSPIRE].

[59] CMS collaboration, *Study of Jet Substructure in pp Collisions at 7 TeV in CMS*, CMS PAS JME-10-013.

[60] Y. Cui, Z. Han and M.D. Schwartz, *W-jet Tagging: Optimizing the Identification of Boosted Hadronically-Decaying W Bosons*, *Phys. Rev.* **D 83** (2011) 074023 [arXiv:1012.2077] [InSPIRE].

[61] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D 80** (2009) 051501 [arXiv:0903.5081] [InSPIRE].

[62] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D 81** (2010) 094023 [arXiv:0912.0033] [InSPIRE].

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

[64] M. Cacciari, S. Frixione, M.L. Mangano, P. Nason and G. Ridolfi, *Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC*, *JHEP* **09** (2008) 127 [arXiv:0804.2800] [InSPIRE].

[65] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *JHEP* **02** (2003) 027 [`hep-ph/0208156`] [ɪɴSPIRE].

[66] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001 [`hep-ph/0206293`] [ɪɴSPIRE].

[67] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J.* **C 53** (2008) 473 [`arXiv:0706.2569`] [ɪɴSPIRE].

[68] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [`hep-ph/0603175`] [ɪɴSPIRE].

[69] M. Cacciari and G.P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett.* **B 641** (2006) 57 [`hep-ph/0512210`] [ɪɴSPIRE].

[70] M. Cacciari, G.P. Salam and G. Soyez, *FastJet*, http://fastjet.fr.

[71] P.-A. Delsart, K.L. Geerlings, J. Huston, B.T. Martin and C.K. Vermilion, *SpartyJet 4.0 User's Manual*, `arXiv:1201.3617` [ɪɴSPIRE].

[72] http://projects.hepforge.org/spartyjet/.

[73] J.R. Christiansen and T. Sjöstrand, *Weak Gauge Boson Radiation in Parton Showers*, *JHEP* **04** (2014) 115 [`arXiv:1401.5238`] [ɪɴSPIRE].

[74] F. Krauss, P. Petrov, M. Schoenherr and M. Spannowsky, *Measuring collinear W emissions inside jets*, *Phys. Rev.* **D 89** (2014) 114006 [`arXiv:1403.4788`] [ɪɴSPIRE].

[75] J. Anderson et al., *Snowmass Energy Frontier Simulations*, `arXiv:1309.1057` [ɪɴSPIRE].

[76] S. Schaetzel and M. Spannowsky, *Tagging highly boosted top quarks*, *Phys. Rev.* **D 89** (2014) 014007 [`arXiv:1308.0540`] [ɪɴSPIRE].