

Tagging a boosted top quark with a τ final state

Amit Chakraborty^{1,*}, Amandip De^{2,†}, Rohini M. Godbole^{2,‡} and Monoranjan Guchait^{3,§}

¹*Department of Physics, School of Engineering and Sciences, SRM University AP, Amaravati 522240, India*

²*Centre for High Energy Physics, Indian Institute of Science, Bangalore 560012, India*

³*Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India*



(Received 4 June 2023; accepted 24 July 2023; published 8 August 2023)

Boosted top quark tagging is one of the challenging, and at the same time exciting, tasks in high energy physics experiments, in particular in the exploration of new physics signals at the LHC. Several techniques have already been developed to tag a boosted top quark in its hadronic decay channel. Recently tagging the same in the semileptonic channel has begun to receive a lot of attention. In the current study, we develop a methodology to tag a boosted top quark ($p_T > 200$ GeV) in its semileptonic decay channel with a τ -lepton in the final state. In this analysis, the constituents of the top fatjet are reclustered using jet substructure technique to obtain the subjets, and then b - and τ -like subjets are identified by applying standard b - and τ -jet identification algorithms. We show that the dominant QCD background can be rejected effectively using several kinematic variables of these subjects, such as energy sharing among the jets, invariant mass, transverse mass, N -subjettiness etc., leading to high signal tagging efficiencies. We further assess possible improvements in the results by employing multivariate analysis techniques. We find that using this proposed top-tagger, a signal efficiency of $\sim 77\%$ against a background efficiency of $\sim 3\%$ can be achieved. We also extend the proposed top-tagger to the case of polarized top quarks by introducing a few additional observables calculated in the rest frame of the $b - \tau$ system. We comment on how the same methodology will be useful for tagging a boosted heavy BSM particle with a b and τ in the final state.

DOI: [10.1103/PhysRevD.108.035011](https://doi.org/10.1103/PhysRevD.108.035011)

I. INTRODUCTION

Top quark [1], the heaviest fermion in the Standard Model (SM), has a large coupling with the Higgs boson. Hence it plays an important role in many of the suggested ideas of beyond the standard model (BSM) physics, as most of these try to address the issue of radiative stability of the Higgs mass. The top being the heaviest, dominates these radiative corrections. Its coupling with the Higgs boson holds the promise of testing the Higgs sector of the SM and beyond (see, e.g., [2–5]) as well as of probing different suggestions/formalisms of BSM physics. Many of these models have a heavy top partner or resonances whose decays involve top quarks and thus provide opportunities to probe these BSM ideas by studying top quark production at the LHC.

Top quarks are of course produced copiously at the hadron colliders, in pairs or singly, respectively, through strong and weak interactions. As mentioned above, another source of t -quarks produced at colliders is the decay of hypothetical heavier particles, predicted in various BSM scenarios, to final states containing the top quark. Examples of such BSM models are supersymmetry (SUSY) [6–10] or little Higgs model [11,12]. Both of these are in fact models which address the hierarchy problem. Models with extra space dimension [13] also predict exotic heavy resonances, which would decay to a final state containing one or more top quarks. In all cases, the decay vertices are likely to carry an imprint of the BSM in their strength and chiral structure. Hence, a study of the production and decay of the top quarks at colliders provides an excellent avenue to explore BSM physics [14,15]. The produced t -quark can be successfully tagged in all its decay modes: pure hadronic as well as the semileptonic ones, where the W in $t \rightarrow bW^+$ decays hadronically or leptonically respectively.

In the context of the study of t -quarks at colliders, knowledge of the polarization of the produced top quark can provide us with an additional important handle to get information on the interaction vertex. This in turn can shed some light on BSM physics responsible. Luckily, the top decay products can be good polarimeters for the decaying quark. This is facilitated by the large mass of the t -quark.

*amit.c@srmmap.edu.in

†amandipde@iisc.ac.in

‡rohini@iisc.ac.in

§guchait@tifr.res.in

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

As a result of this large mass, the t -quark, with a lifetime of $\sim 5 \times 10^{-25}$ s, decays before its hadronization which occurs on a timescale of $1/\lambda_{\text{QCD}} \approx 10^{-24}$ s. As a result, its decay products can carry information about its spin state. In fact, it has been long known that the decay product angular distributions with respect to the spin direction of the decaying top as well as their energy distributions depend on the t -polarization [16–18] and hence can be used to gain information about the same. Since t -quark polarization is an important probe of BSM physics and the aforementioned correlations follow from the chiral structure of the SM tbW vertex, the effect of anomalous tbW couplings on these correlations also has to be investigated for them to be useful probes of polarization. Such investigations [19–29] have shown that the angular distributions of the down type fermion (ℓ, d), in the decay of the W coming from the top, are particularly robust probes of the t -polarization.

With the ever-increasing lower limits on the masses of the BSM particles, the top quarks, expected to be produced in their decays, will necessarily have higher transverse momenta and hence will be highly boosted. This large boost causes the decay products to be highly collimated and these appear in the detector as a single jet in both the hadronic and semileptonic decay modes. Tagging these boosted top quarks at the LHC has been an active field of research now for more than a decade.

The opening angle between the decay products depends inversely on the top decay Lorentz factor $\gamma \sim E/m$. The jet which includes all the decay products of the decaying boosted top quark tends to have a larger radius than a typical QCD jet which owes its structure to the light parton radiations. The jet corresponding to a boosted top quark is thus a “large- R jet” [30]. A number of tagging algorithms have been proposed corresponding to hadronic top quark decays using large- R jet analysis. These range from those based on substructure [31–33] to the recent ones which include state-of-the-art deep learning methods [34–38]. The jet substructure based taggers make use of the identification of W -boson and top quark through the mass reconstruction while working with jet constituents. The strategies based on jet images have been the main thrust in the deep learning algorithms (for more details, see Refs. [39,40]).

Tagging of a boosted top quark decaying semileptonically referred to as leptonic top jet suffers due to the presence of the neutrinos in the final state, which hinders complete reconstruction of the top quark mass. One can of course resort to the transverse mass variables such as M_T and extract the mass of the decaying top quark from the edge of the M_T distribution. Alternatively, it is also possible to use the presence of hard tracks originating from the leptons inside a large- R jet, to tag [37] a leptonically decaying boosted top quark. The authors in [41] have shown that it is possible to construct kinematic quantities, which can discriminate between a boosted leptonic top-jet

containing non-isolated electrons/muons and a QCD large- R jet where light jets are mistagged as leptons.

In this work, we devise a tagging method to identify boosted top quarks decaying semileptonically with a τ -lepton in the final state, taking into account decays of the τ both in its leptonic and hadronic channels. In addition to the ν_τ from the W decay, the final state contains one more neutrino coming from the τ decay as well. This complicates the tagging process. It should be noted that, since the τ -leptons are heavier than other leptons, their couplings and hence polarization are also sensitive to new physics effects. Hence having a top-tagger for a t decaying with a τ in the final state can open up further possibilities of BSM studies using the polarization of the τ [42–45] as well. The major challenge in identifying a τ -jet (τ_h), which arises due to the hadronic decay mode of the τ inside a large- R jet, is to distinguish it from quark and gluon-initiated QCD jets. The proposed top tagger relies on efficiently identifying an energetic b -jet and a τ -jet within the top-candidate jet. We benchmark the performance of our proposal by using simulated events, corresponding to the production of a heavy W' followed by its decay $W' \rightarrow tb$ and further the decay of the t into a $b\tau\nu_\tau$ final state, applying the jet substructure technique and constructing a few discriminating kinematic observables. We demonstrate that these are very useful in eliminating QCD jets faking as τ -jets and thus facilitate tagging the top jet. We achieve an efficiency of $\sim 77\%$ for tagging the semileptonic top quark jet with taus in the final state, while keeping the mistagging efficiencies of backgrounds from light flavor QCD jets to $\sim 3\%$ level.

Even though the main focus of this study is to identify top quark jets with τ -leptons in the final state, the proposed methodology can be applied to any large- R jet, which includes a b quark and a τ . For example, a light charged Higgs boson with $m_{H^\pm} < m_{\text{top}}$, is an example. In this case, the top quark can decay to a bottom quark and a charged Higgs boson which then subsequently dominantly decays through $\tau\nu_\tau$ mode. Our proposed methodology can also be used to probe decays of the third generation squark, namely the top squark, in R -parity violating (RPV) scenarios. In RPV SUSY model, both bilinear (LH) and trilinear (LLE, LQD) renormalizable lepton number violating operators are allowed in the superpotential [9,46] by gauge invariance and supersymmetry. In this scenario, RPV decay of top squark $\tilde{t} \rightarrow b\tau$, can occur for both the bilinear and trilinear RPV terms. See, for example, [47] and [48], respectively. Clearly, our methodology can be used to tag the top squark decaying in this fashion. A third generation leptoquark of electric charge of $+4/3$ unit can also have decays to a final state containing a b and a τ similar to the top quark jet. In these cases, the absence of neutrinos from W decay implies that, unlike the top quark mass, it is possible to reconstruct the mass of Leptoquark or top squark in a straightforward way once we tag the b - and τ -subjects efficiently.

This paper is organized as follows. In Sec. II, we introduce the methodology used for tagging top quarks with taus in the final state and discuss the identification strategy for the b - and τ -subjets which are the first steps in this study. In Sec. III we introduce several variables which have the power to discriminate the signal from the background. The results of the multivariate analysis are presented in Sec. IV. In Sec. V, we construct a few polarization sensitive observables to explore the ability of our top-tagger to differentiate between left- and right-polarized top quarks. Finally, we summarize in Sec. VI.

II. METHODOLOGY: b AND τ -JET IDENTIFICATION

The challenging part of tagging a boosted top quark in the $t \rightarrow b\tau^+\nu_\tau$ decay mode is to identify the b - and τ -like subjets inside the top-candidate jet. In this section, we describe this strategy very systematically. For the purpose of simulation of boosted top quark signal, we generate events for the production of a heavy W' boson with its subsequent decay $W' \rightarrow tb$ and further the decay of the t into a $b\tau^+\nu_\tau$ final state. We do this for the LHC center of mass energy $\sqrt{s} = 13$ TeV. Henceforth, we refer to these as W' events. The generated process is indicated in Eq. (2.1). Even though the Eq. (2.1) and the Feynman diagram correspond only to W'^+ , we have, of course, generated events for W'^- as well.

$$pp \rightarrow W'^+ \rightarrow t\bar{b} \quad \downarrow \quad \tau^+\nu_\tau b, \quad (2.1)$$

and the corresponding leading order (LO) Feynman diagram is shown in Fig. 1.

We consider the W' effective model [49] to generate the signal events setting $m_{W'} = 1$ TeV. This model is an extension of the Standard Model (SM) incorporating a W' boson with arbitrary vector and axial-vector couplings to the SM quarks. Following [49], the relevant part of the Lagrangian can be written as,

$$\mathcal{L} = \frac{g}{\sqrt{2}} V_{f_i f_j}^{\text{CKM}} \bar{f}_i \gamma_\mu (k_R P_R + k_L P_L) W' f_j + \text{H.c.}, \quad (2.2)$$

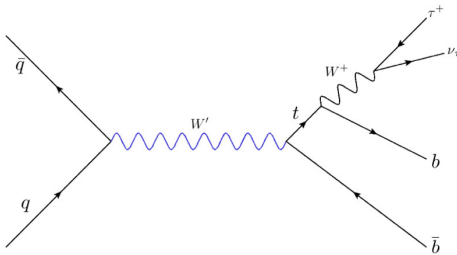


FIG. 1. Feynman diagram at leading order for the signal process $pp \rightarrow W'^+ \rightarrow t\bar{b} \rightarrow \tau^+\nu_\tau b\bar{b}$.

where $k_R(k_L)$ are the right-handed (left-handed) W' boson gauge couplings to quarks f_i and f_j , $V_{f_i f_j}^{\text{CKM}}$ are the CKM matrix elements, and $P_{R,L} = (1 \pm \gamma_5)/2$, g is the SM $SU(2)_L$ coupling. For a SM W boson, $k_R = 0$, $k_L = 1$. Relative values of k_L and k_R determine the polarization of the produced top quark. Its value \mathcal{P}_0 in the rest frame of W' can be calculated easily by computing Γ^\pm , the partial decay widths of the $W' \rightarrow t_\lambda \bar{b}$ with helicity $\lambda = \pm$ and is given by,

$$\mathcal{P}_0 = \frac{\Gamma^+ - \Gamma^-}{\Gamma^+ + \Gamma^-} = \frac{(k_R^2 - k_L^2)(2 - x_t^2 + x_b^2) \sqrt{1 + (x_t^2 - x_b^2)^2 - 2(x_t^2 + x_b^2)}}{12k_R k_L x_t x_b + (k_R^2 + k_L^2)(2 - (x_t^2 - x_b^2)^2 - (x_t^2 + x_b^2))}, \quad (2.3)$$

with $x_t = m_t/m_{W'}$, $x_b = m_b/m_{W'}$. If we were to neglect the t , b masses, then the produced t will always be left-handed for $k_L = 1$, $k_R = 0$ and right-handed for $k_L = 0$, $k_R = 1$. If we use Eq. (2.3) to calculate \mathcal{P}_0 , for $m_{W'} = 1$ TeV, $m_t = 172$ GeV, and $m_b = 4.7$ GeV, and $k_R = 1$, $k_L = 0$ and $k_R = 0$, $k_L = 1$ we get $\pm 0.97 (\sim \pm 1)$, respectively. Thus the large mass of the W' implies that the polarization of the produced t -quark will be decided completely by values of k_L and k_R . To develop the tagging methodology, we generate unpolarized boosted top quarks from W' decay by setting $k_L = k_R$. For the polarization study, we set k_R (k_L) to zero to produce left (right) -polarized top quarks.

The W' events ($pp \rightarrow W'^+ \rightarrow t\bar{b}$) are generated using MadGraph5_aMC@NLO [50]. It is necessary to take into account the effect of the spin correlations as well as the finite width of the decaying top quark in the decay $t \rightarrow b\tau^+\nu_\tau$. We do this by employing the MadSpin [51] method. Next, W' events are passed through PYTHIA8 [52] for parton shower and hadronization. In order to consider detector effects, those events are passed through DELPHES v3.4 [53] with compact muon solenoid (CMS) card setting. The process of identifying boosted top jets begins with the construction of large- R jets, setting jet radius parameter $R = 1.5$ and anti- k_T jet clustering algorithm [54] as implemented in FastJet v3.2.1 [55]. Then, we select those large- R jets which pass the threshold of transverse momentum, $p_{T_j}^{\text{min}} = 200$ GeV. The boosted large- R jets are contaminated by several sources, such as soft radiation, underlying events and multi-particle interactions. These are removed by applying SoftDrop technique [56] setting free parameters β and z_{cut} to their values for the standard CMS choice [57], viz. $\beta = 0$ and $z_{\text{cut}} = 0.1$. The constituents of the soft-dropped large- R jet are further re-clustered with jet radius parameter $R = 0.5$ using anti- k_T algorithm to form subjets with a minimum $p_{T_j}^{\text{min}} = 20$ GeV. In principle, W' is produced at almost rest; hence W' decay to b and t will produce two back-to-back

large- R jets. For the above mentioned choices of various parameters, we find that for $m_{W'} = 1$ TeV about 58% events contain two back-to-back large- R jets whereas the fraction of these events rises to $\sim 79\%$ for $m_{W'} = 3$ TeV. To ensure that a large- R jet is indeed a top jet, it is necessary to identify the subjets in it as b - and τ -jets. We will describe our strategy for identification of the b - and τ -like subjets after describing the different event samples we generate for the backgrounds as well.

High p_T QCD jets can fake as top jets. Hence, we need to study the impact of jets produced via QCD processes while developing the strategy to tag top jets in the semileptonic channel. The kinematics of the top decay looks similar to QCD parton splitting when its boost factor y_t is $\sim 1/\alpha_s$. Therefore, at high p_T , the subjets of the QCD large- R jet, are very likely will be misidentified as b - or τ -jets. Hence, efficient identification of the subjets as b - and τ -jet is necessary to suppress the number of QCD events significantly. Various properties of the QCD jets differ significantly from the top-candidate jet and the idea is to exploit these differences to reduce the QCD contamination. One such property is the average mass of the jet, which is affected mainly by the sharing of energy among different members of the jet. For a QCD large- R jet this average mass increases with the p_T of jets. It has been shown that for QCD jets with p_T more than ~ 300 GeV, the corresponding mean jet invariant masses lie within the window of 30–160 GeV [31]. This mass range, covering the semileptonic top mass, is precisely the one that is important for our analysis which aims to tag a top quark.

QCD jets have a steeply falling p_T distribution, but due to the much larger production cross-section of the jets, there is a significant number of jets even after the requirement of large minimum p_T of 200 GeV for the jet. The steeply falling p_T spectrum of these jets means that one has to take extra care to generate appropriately large number of events. QCD events from $pp \rightarrow jj$ are generated in three p_T^{bins} : [200–300, 300–600, > 600] GeV where p_T corresponds to the transverse momentum of hard scattered particles in the final state. The number of simulated events is determined by keeping in mind that we perform our analysis for a luminosity of 10 fb^{-1} and we need to pay particular attention to the region [300–600] GeV, since the QCD jets in this p_T^{bin} contribute dominantly to a window in jet mass which is populated by the signal from semileptonic decay of the top quark. We have simulated 10, 7, and 3 M events in these three p_T^{bins} , respectively. We also need to simulate the hadronically decaying top-antitop quark pair ($t_h \bar{t}_h$) events which is a potential background. The list of generated events used for simulation is presented in Table I along with the range of p_T^{bins} for the dijet QCD events as well as for the $t_h \bar{t}_h$ events. Next, we proceed to discuss the identification of the subjets as τ - and b - subjets, which, as mentioned above, is very important so as to be able to handle the background from QCD jets.

TABLE I. List of the signal and background samples used with their kinematics.

Process	p_T^{bins} (GeV)	Cross section (pb)
Dijet QCD	200–300	$\approx 4.9 \times 10^4$
	300–600	$\approx 7.9 \times 10^3$
	> 600	$\approx 2.0 \times 10^2$
$t\bar{t}$ Hadronic($t_h \bar{t}_h$)	> 150 GeV	≈ 208.2
$W' \rightarrow t\bar{b} \rightarrow \tau b\bar{b}\nu$	Mass of $W' = 1$ TeV	≈ 0.6

(i) τ -jet identification:

The τ -lepton decays hadronically with a probability of 65% producing charged (mainly π^\pm) and neutral hadrons (π^0). Hence the multiplicity of decay products, particularly charged tracks, is low in numbers and they are highly collimated in a cone with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 1.5$, where $\Delta\eta$ and $\Delta\phi$ are the differences of pseudorapidities and azimuthal angles respectively between two particles. In collinear approximation, i.e. if $p_T^\tau \gg m_\tau$, the decay products are collimated even more so as to be contained in a jet of even smaller radius, say $R < 0.5$. Among the τ decay products, the neutral pions deposit a considerable fraction of electromagnetic energy in the calorimeters through photon. This is accompanied by one or three-prong low p_T charged track multiplicity observed in the tracker, which are the characteristics of a τ -jet. Currently, ATLAS and CMS have developed very sophisticatedly dedicated algorithms for the identification of τ -jets using attributes of τ decay, such as energy difference in calorimeter cells, lifetime and mass, track multiplicities etc. [58–60]. However, those techniques are beyond the scope of our present analysis. Instead, we use a naive track-based isolation algorithm to identify τ -like subjets.

In this procedure [58]; first, we identify the tracks which are within the jet-track matching cone with radius $R_m = 0.1$ calculated using the candidate τ -jet axis, and then select those tracks with minimum p_T of 2 GeV (see Fig. 2). Among these tracks, we identify the leading track (or the seed track) with minimum $p_T > 6$ GeV and $|\eta| < 2.5$. Other tracks are accumulated in a narrow region around the seed track with a cone radius $R_s = 0.07$, and we demand that the difference of the z -impact parameter Δz_{tr} between the leading track and these selected tracks (signal cone) is smaller than 2 mm. This additional requirement ensures that all the tracks in the signal cone are coming from the same τ -lepton decay. Furthermore, we adopt a track isolation method, where isolated tracks are reconstructed in a larger

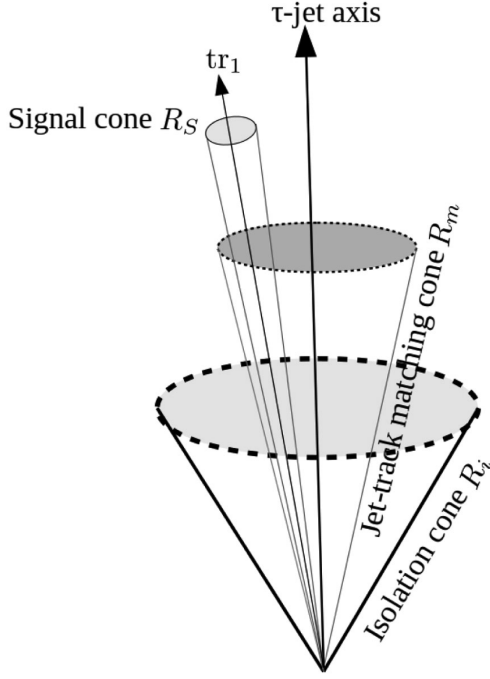


FIG. 2. Basic principle of τ -jet identification with charged track isolation.

cone size $R_i = 0.45$ around the candidate τ jet axis with minimum transverse momentum p_T^i of 1 GeV. The isolation criterion is satisfied when the number of tracks (1 or 3) in the isolation region is the same as in the signal cone. Following this naive technique, we achieve τ -identification efficiency $\epsilon_\tau \sim 60\%$ for a moderate range of $p_T = 20\text{--}60$ GeV of τ -jets. In case of QCD, the misidentification efficiency of τ -jets is $\sim 5\text{--}6\%$.

The effectiveness of the τ identification method is demonstrated in Fig. 3 in distributions of multiplicity

of the charged tracks ($N_{\text{trk}}(\tau_h)$), mass (m_{τ_h}) for the identified τ -jets. The left plot of charge track multiplicity shows a clear tendency to one and three prong structures for identified τ -jets, as expected. QCD jets acquire mass through multiple splitting, and the EM clusters are far away from the jet axis than that of signal; therefore, a smeared distribution is more likely. While a clear peak structure is visible in the mass distribution for the signal events, shown by the solid blue line (right plot in Fig. 3), the invariant mass distribution for the misidentified τ -jet shows a long tail, for the QCD jets as well as the jets from $t_h\bar{t}_h$, displayed in the same plot by red dotted and black dash dotted lines respectively. It is true that the invariant mass distribution for the signal events does have a somewhat long tail. However, it is possible to reduce it further by using isolation criteria in addition to the single one that we have used [61].

(ii) b -jet identification:

A subjet inside the candidate top large- R jet is identified as a b -jet, if the angular distance ΔR between the jet and the nearest B -hadron satisfy $\Delta R < 0.5$. For signal events, B -hadrons dominantly arise from the b -quarks, which are produced through the decay of W' and the t . Intuitively, the B hadrons, which come from the b quark originating from the top quark decay, are likely to satisfy the above matching condition viz. $\Delta R < 0.5$. On the other hand, for QCD multijet events, B hadrons mostly originate due to a gluon splitting into a $b\bar{b}$ pair and hence unlikely to be close to the large- R jet axis. In b -jet identification strategy, we also take into account the impact of detector effects by incorporating the b -tagging efficiencies and mistag rates reported in [62] and which are summarized in Table II.

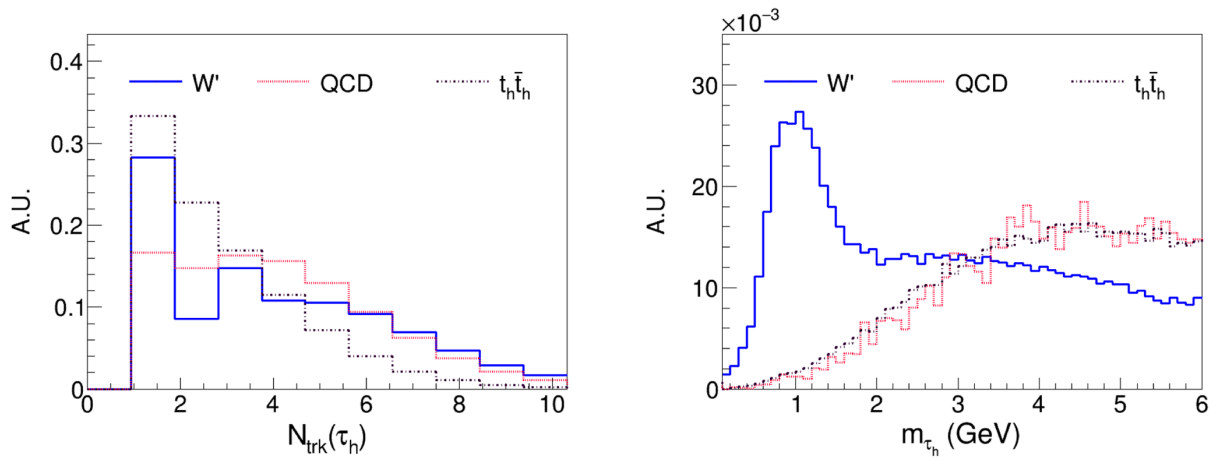


FIG. 3. Distribution of the number of charged tracks inside the τ -subjet (left) and mass of the τ -subjet (right) for signal (W') and background events. The solid blue line shows the distribution for the signal while the same for QCD and $t_h\bar{t}_h$ are shown in red dotted and black dash dotted lines, respectively.

TABLE II. b -tagging efficiencies for different p_T range of the jets following the performance of ATLAS b -identification algorithms with Run2 data [62].

p_T of the b -jet (GeV)	b tagging efficiency (%)
Upto 60	60
[60,200]	80
[200,400]	60
400–	50

Following the CMS analysis, we have used a mistag rate to be 2% for a light jet to be identified as a b -jet, irrespective of p_T of the jet [63]. In our simulation, we correct the b identification probability by applying all these efficiencies. We get $\sim 77\%$ identification efficiency for b .

Figure 4 shows distributions in the invariant mass of the large- R jets containing b - and τ -like subjets, identified to be so, using the above procedure for all the three types of events (Table I) that have been generated, viz. W' and hadronic top and QCD events. For the sake of comparison, the same invariant mass constructed out of the total four-momentum of b -quark and the visible decay products of the τ -lepton (τ^{vis}), referred to as $b - \tau^{\text{vis}}$ system is shown. At the truth level, mass of $b - \tau^{\text{vis}}$ system $m_{b\tau^{\text{vis}}}$ is expected to be bounded by $m_{b\tau^{\text{vis}}} < \sqrt{m_t^2 - m_W^2} \approx 152$ GeV in the limit $m_b \rightarrow 0$. A clear peak is visible in the distribution for parton level events around ~ 75 GeV. The reconstructed top jet mass distribution from W' event is found to have a peak too around the same value, and the distribution is smeared due to the hadronization of the b -quark and consequent decay of the B -meson as well as the decay of the τ lepton and the detector effects. Notice that the

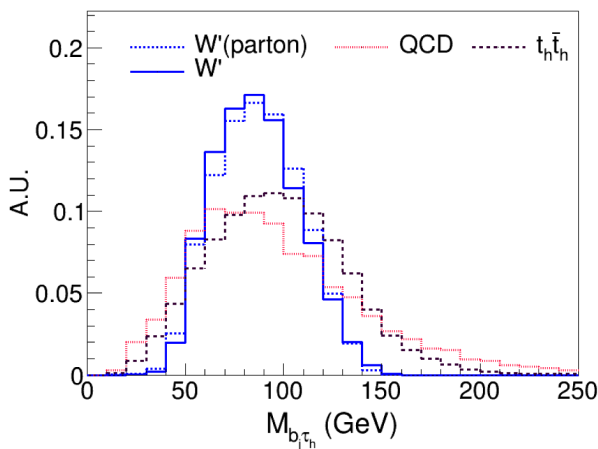


FIG. 4. Distribution of invariant mass of $b - \tau$ jet system for signal and background events. The parton level (blue dotted line) curve for W' is constructed out of momenta of the b -quark and visible decay products of τ . The color code for all the other curves is the same as in Fig. 3.

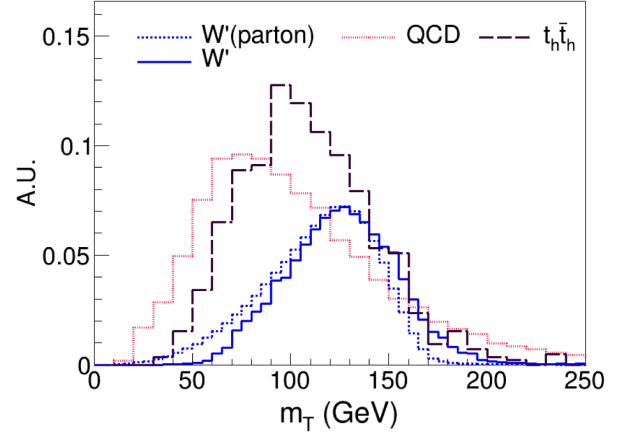


FIG. 5. Distribution of the transverse mass constructed out of b -subjet (b_j), τ -subjet (τ_h), and MET for the signal and background events. The color code is the same as in Fig. 4.

distribution does not show any peak-like structure for the $t_h\bar{t}_h$ and QCD events. This indicates to us that a window in the distribution of invariant mass of the b - and τ -subjets, i.e., $M_{b_j\tau_h}$, say between 60 to 160 GeV, could be chosen optimally to tag a semileptonic t -jet.

III. TOP JET IDENTIFICATION

In this section, we construct a number of observables that can be used to identify top jets. We do this by exploiting the features of the b - and τ -like subjets that we observed in the earlier section.

(i) Transverse mass:

The dominant source of missing transverse energy (MET) in the signal is due to the presence of two neutrinos: one from the decay of W and the other from the decay of the τ as can be seen from Eq. (2.1). The presence of two neutrinos makes it difficult to reconstruct the mass of the top large- R jet, compared to the case of the hadronically decaying t , where such reconstruction plays an important role in its tagging. However, the direction of MET can help to do the job in the present case because of the boosted nature of the t and the W . In fact, as a result of the boosted nature of the t , W , and τ , the direction of MET is expected to be collinear to τ decay products. This feature can be utilized to get an additional handle to identify top jets and also to reduce background. Keeping the above kinematics in mind, we require the ΔR between MET and the candidate top jet to be < 3 . We construct a transverse mass observable m_T , combining the momenta of the identified b - and τ -jet, along with the MET given by,

$$m_T^2 = M_{b_j\tau_h}^2 + 2(E_T^{b_j\tau_h} \cancel{E}_T - \mathbf{p}_T^{b_j\tau_h} \cdot \cancel{\mathbf{p}}_T), \quad (3.1)$$

where $M_{b_j\tau_h}$ and $\mathbf{p}_T^{b_j\tau_h}$ denote the invariant mass and transverse momentum of the $b - \tau$ jet system respectively. At parton level we again use the $b - \tau^{\text{vis}}$ system as discussed previously and define the m_T as, $m_T \equiv \sqrt{m_{b\tau^{\text{vis}}}^2 + 2(E_T^{b\tau^{\text{vis}}} E_T^{\text{gen}} - \mathbf{p}_T^{b\tau^{\text{vis}}} \cdot \mathbf{p}_T^{\text{gen}})}$. The m_T of $b - \tau$ jet, displayed by the solid blue curve in Fig. 5 is expected to have an endpoint at t mass, which is smeared due to the difference in correlation of MET coming from two neutrinos with the top jet as well as other effects mentioned in the previous section. We can use this feature of the m_T distribution to select the candidate large- R jet after requiring an invariant mass of $M_{b_j\tau_h}$ in the aforementioned mass window of [60–160] GeV, which removes a good fraction of background events.

(ii) Energy sharing of subjects:

The pattern of energy sharing of subjects is very different for the signal and backgrounds. This facilitates the construction of an observable which offers good separation between boosted top jets from the signal and the background. The fraction of energy carried by a subject (j) of a large- R jet (J) as

$$Z_j = \frac{E_j}{E_J}. \quad (3.2)$$

In Fig. 6, we show the Z_j distribution for τ_h (left panel) and b -like(right panel) subjects corresponding to both signal and background events. For one-to-one correspondence with the jet level, the energy fractions at the parton level are defined as the ratio of the energy of the b quark or energy of the visible decay products of the τ to the sum of the two ($E_b + E_{\tau^{\text{vis}}}$). For signal events, it is expected that the b -subject will carry a large fraction of energy while τ_h will share

comparatively a smaller fraction of energy of the top system since a fraction is taken away by MET which contains the neutrino from the τ decay. One can see that Z_{τ_h} peaks around 0.2. However, for background events, these energy fractions of the b - and τ - like jets are uncorrelated. In fact, Z_{b_j} in QCD, the dotted red line in the right plot of Fig. 6, shows a flat distribution, whereas Z_{τ_h} for a QCD jet mimicking a τ - like subject is dominantly distributed to much lower values. The peculiar sharing of energies, particularly for Z_{b_j} in QCD events, can be attributed to additional components the subject contains due to soft radiation, which could not be removed even after the application of soft drop method [56].

For the QCD background, due to the presence of multiple soft emissions, the total energy of the subjects is not shared only among the misidentified b - and τ -jet. Hence, the behavior of the energy fractions is different for the signal and the background. Further, these multiple [soft subjects make a non-negligible contribution to the mass of the large- R jet. We construct a new variable $\Delta X_{b_j\tau_h}$, which reflects the impact of the presence of the multiple soft subjects in the large- R jets for the QCD background. This variable $\Delta X_{b_j\tau_h}$ is the deviation of the ratio of the mass of the b - and τ - subject system ($M_{b_j\tau_h}$) with respect to the mass of the corresponding identified candidate top jet (M_J) from 1,

$$\Delta X_{b_j\tau_h} = 1 - \frac{M_{b_j\tau_h}}{M_J}. \quad (3.3)$$

Clearly, for signal events, it is expected to peak around $\Delta X_{b_j\tau_h} \sim 0$ as shown in the blue line of Fig. 7. For the backgrounds, as explained above, b -jets are contaminated by soft components and hence are

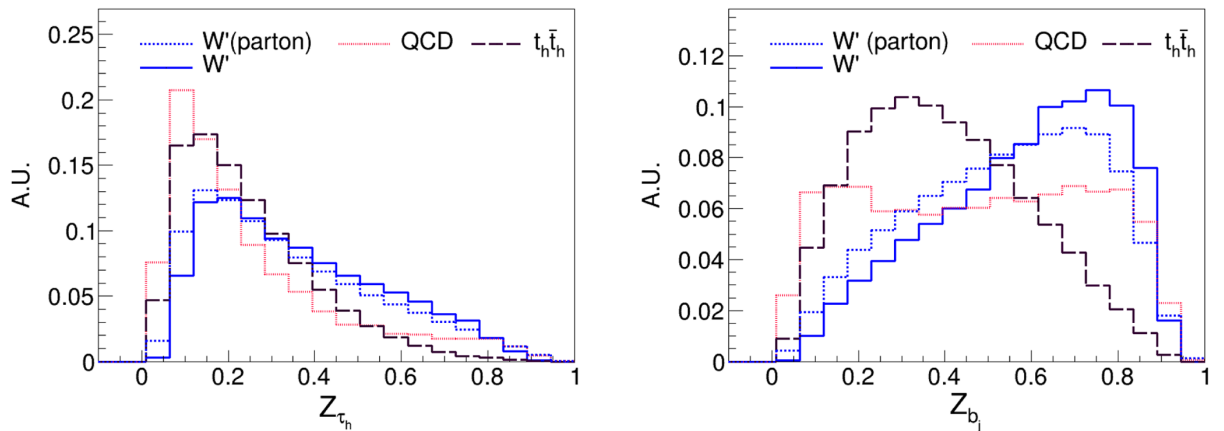


FIG. 6. Distribution of energy fraction of τ_h (left) and b_j (right) in top jet for signal and background events. The color code is same as Fig. 4.

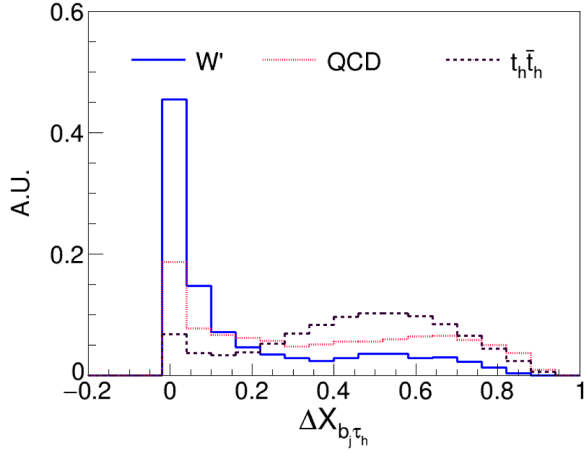


FIG. 7. Distribution of variable $\Delta X_{b_j \tau_h}$ [Eq. (3.3)] for signal and background events. Color code is the same as in Fig. 3.

much flatter and extend to larger values. We find that this variable helps improve the top tagging efficiency.

(iii) N -subjettiness:

Another very useful variable in the jet substructure technique, which helps to improve our study is N -subjettiness. Subjettiness [64] takes advantage of the different energy flow in the different particles present within the large- R jet. It effectively counts the number of subjects in a given jet. If there are N candidate subjects in a specific jet, one calculates subjettiness as,

$$\tau_N = \frac{1}{\sum_k p_{T,k} R_{\text{Jet}}} \sum_k p_{T,k} \min \times \{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}. \quad (3.4)$$

Here k runs over all the constituents of a jet of momentum $p_{T,k}$ and $\Delta R_{J,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is

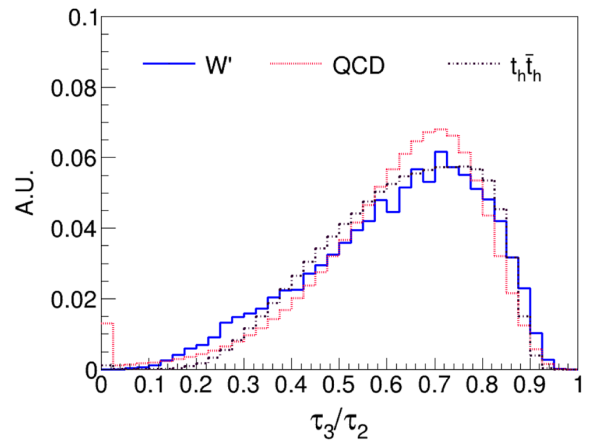
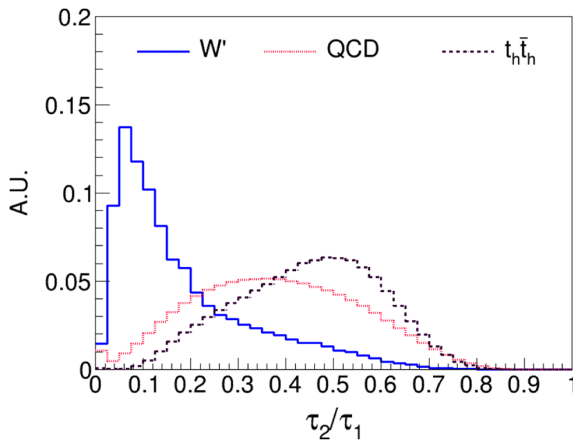


FIG. 8. Distribution of τ_2/τ_1 (left) and τ_3/τ_2 (right) for leptonic top and background jets. Discriminators like τ_2/τ_1 measure the relative alignment of the jet energy along the individual subject directions. Color code is the same as Fig. 3.

the distance in $\eta - \phi$ plane between a candidate subject j and a constituent k . The normalization factor is taken to be the jet p_T multiplied by its radius. In the limit $\tau_N \rightarrow 0$, the jet must have $\Delta R_{N,k} = 0$, i.e., all the radiation is perfectly aligned along the candidate subjects, and therefore the jet has exactly N subjects. In case of $\tau_N \rightarrow 1$, the jet must have a large fraction of its energy distributed away from the candidate subject direction, therefore it has at least $N + 1$ subjects, i.e., the minimization missed some subject axes.

Therefore, jets with smaller τ_N are said to be N -subjettiness, whereas larger τ_N have more than N subjects. QCD jets can have larger values of subjettiness variables due to diffuse spray of large angle radiation, and hence individually τ_1 or τ_2 cannot provide much distinction between signal and background events. The ratio τ_2/τ_1 , however, is a different story and can be effective in discriminating the different two-prong objects. Similarly, τ_3/τ_2 is a better choice while probing three-prong objects. We show the distribution of these two variables viz., τ_2/τ_1 and τ_3/τ_2 in the left and right panel of Fig. 8 respectively. For the leptonically decaying t -quark the large- R jet is essentially a two prong object and hence τ_2/τ_1 is found to be a good discriminating observable.

Based on the various observations in the discussion above we list the steps to be followed systematically to identify a boosted top jet in its decay channel, $t \rightarrow \bar{b}\tau^+\nu_\tau$.

- (1) Cluster the final state hadrons of the events to jets with a minimum p_T , say $p_T \sim 200$ GeV or larger, setting the jet size parameter $R = 1.5$. Then apply a jet grooming technique (e.g., SoftDrop) and remove the soft and wide angle radiation contamination. Select the large- R jet mass with a window of the range 60–160 GeV. It is likely to be a top candidate jet.

- (2) Recluster the constituents of the candidate fat top jet to a smaller radius ($R \sim 0.5$), and select those events where the candidate jet consists of at least two subjets.
- (3) Identify b - and τ - like subjets following the procedure as described in Sec. II.
- (4) Finally, if at least two different subjets in close proximity originating from the candidate jet pass the b and τ jet identification, the large- R jet can be considered as a top-large- R jet if the invariant mass lies in the window 60–160 GeV.
- (5) Furthermore, construct the following discriminating observables to reduce possible backgrounds.
 - (a) Transverse mass (m_T), constructed from the b – τ subjet system and MET following Eq. (3.1).
 - (b) Energy fractions of the b - and τ - subjets inside the top large- R jet. One can then utilize the difference in the template of energy sharing between the subjets (Z_{b_j}, Z_{τ_h}).
 - (c) The fraction of mass carried by the b – τ jet, Eq. (3.3) ($\Delta X_{b_j\tau_h}$). In defining this observable one has used the excess of softer contamination to subjets in backgrounds as compared to the signal.
 - (d) N -subjettiness variable such as, τ_2/τ_1 and τ_3/τ_2 etc. [cf. Eq. (3.4)].

Figure 9 shows the efficiency ϵ_t of top jet identification obtained after following the above procedure (steps 1–4) for a moderate range of boosted top jet p_T of 200–450 GeV. Here ϵ_t is defined as the ratio of number of candidate large- R jet which has b - and τ - subjets, to the corresponding total number of candidate top jets within matching cone of 1.5 of t -quark, running on all large- R jets in an event. The corresponding top like jet misidentification efficiency for QCD jets is $\sim 1\%$ – 2% . An important observation is that ϵ_t decreases with the increasing p_T of the top jet. It is due to

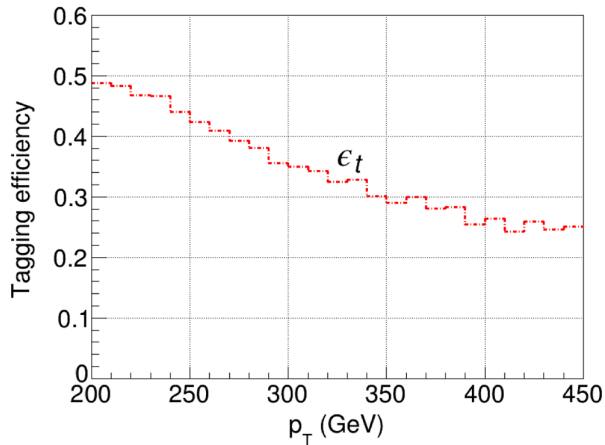


FIG. 9. The efficiencies for identification of the top jet (ϵ_t) in the p_T range of 200–450 GeV.

the fact that for large values of p_T for the large- R jet, the subjets are no longer distinguishable. The tagging efficiency also depends on the radius of the large- R jet. We have checked that by changing ΔR from 1.5 to 1.0 and the subjet radius from 0.5 to 0.3, the tagging efficiency increases by 5%–6%.

We compute the efficiency of rejecting the backgrounds with the following set of cuts as described in Eq. (3.5) on the discriminators mentioned in last step above.

$$\begin{aligned}
 50 < m_T < 200, & \quad 0.3 < Z_{b_j} < 0.9, \\
 0.15 < Z_{\tau_h} < 0.9, & \quad 0 < \tau_2/\tau_1 < 0.45, \\
 0.2 < \tau_3/\tau_2 < 0.9, & \quad \Delta X_{b_j\tau_h} < 0.3
 \end{aligned} \tag{3.5}$$

With the variables mentioned in Eq. (3.5), the signal efficiency integrated over the p_T range of t -jet turns out to be 63.5% where the corresponding mistagging efficiency due for QCD jets is 3.1% and for $t_h\bar{t}_h$ it is 5.9%.

IV. MVA ANALYSIS: UNPOLARIZED TOP

The standard cut-based strategy often rejects a significant fraction of signal events while reducing the background events that mimic the signal events. This situation can be improved further by employing a multivariate analysis (MVA) [65] technique which increases the background rejection rates.

We use the boosted decision tree (BDT) method for the optimization purpose with the TMVA framework [65]. A decision tree takes a set of input features and splits input data recursively based on those features to classify events as either signal-like or backgroundlike. To increase the stability in the training sample with respect to statistical fluctuations, we adopt the *Adaboost* [66] technique. The decision trees are constructed using half of the signal and background events, while the other half is utilized to test the performance of the trained model. We choose the relevant hyperparameters of this method as follows: Number of trees $N_{\text{Tree}} = 850$, maximum depth of the decision tree $\text{MaxDepth} = 5$, and minimum percentage of training events in each leaf node is given by $\text{MinNodeSize} = 2.5\%$; other parameters are set to its default values [66].

Events are selected when the candidate large- R jet includes a b - and a τ -identified subjet inside it. A number of kinematic variables are constructed out of the momenta of these objects, as discussed in the previous section, and eventually 10 input variables are used for BDT training. In Table III, the set of input variables are shown, ranking them according to the importance in the BDT analysis for $m_{W'} = 1$ TeV at $\sqrt{s} = 13$ TeV. The importance here means the effectiveness of those variables in suppressing backgrounds while maintaining better signal purity.

In a typical BDT analysis, a few things need to be taken care of, such as instability, bias and overtraining. To that

TABLE III. List of the variables used to train W' signal, QCD and $t_h\bar{t}_h$ events.

Rank	Variable	Description
1	m_T	Transverse mass of $b - \tau$ jets and MET of the system
2	M_{b,τ_h}	large- R jet Mass distribution in the mass range [60,160] GeV
3	$\Delta X_{b,\tau_h}$	1-Invariant mass of $b - \tau$ jet/large- R jet mass
4	τ_2/τ_1	Ratio of subjetiness of the Soft dropped Top-jet τ_2/τ_1
5	Z_{b_j}	Fraction of energy carried by the identified b -jet of Soft-Dropped Top large- R jet
6	Z_{τ_h}	Fraction of energy carried by the identified τ -jet of Soft-Dropped Top large- R jet
7	τ_3/τ_2	Ratio of subjetiness of the Soft dropped Top-jet τ_3/τ_2
8	m_{b_j}	Mass of identified b - jet
9	m_{τ_j}	Mass of identified τ - jet
10	$N_{\text{trk}}(\tau_h)$	Charged track multiplicity of identified τ - jet

end first and foremost one has to ensure that a sufficient number of events for both signal and backgrounds are generated such that the importance or ranking of the variables is stabilized. To remove any bias coming from a particular QCD process because of the kinematical features of the background, we optimize the number of generated events. For instance, QCD events for low p_T bins, such as 200–300 GeV regime, 10M events are simulated, while for 300–600 GeV and for > 600 GeV regime, approximately 7 M and 3 M events are generated, respectively. We have simulated 1.3M signal events for W' mass = 1 TeV. In order to verify that there is no overtraining of the trees, the sample size of both training and testing datasets are optimized so that no deviation is observed in the final outcome. The goodness of fit is also checked with the Kolmogorov-Smirnov (KS) test, and observed that the KS value is within the permissible range of [0,1] and closer to the mean 0.5.

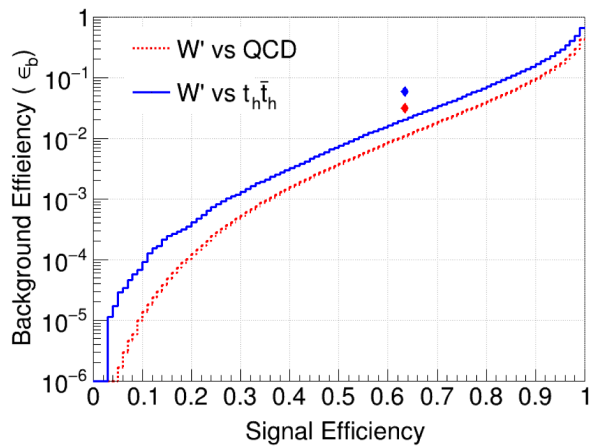


FIG. 10. The signal and background efficiencies for $m_{W'} = 1$ TeV against QCD (red dotted) and $t\bar{t}$ (blue solid). The two points in red and blue color represent the corresponding efficiencies for QCD and $t_h\bar{t}_h$ from a cut based analysis with the choice of cuts mentioned in Eq. (3.5).

After the classifier has been trained, it gives the output in terms of a single variable, the BDT response. Applying a cut on the BDT output variable, the signal-to-background ratio is optimized and can be presented as the receiver operative characteristic (ROC). In Fig. 10, we display the ROC for our proposed tagging technique. The figure estimates the tagger performance at different background rejection rates ($1 - \epsilon_b$), where ϵ_b is the background acceptance efficiency. For this classifier the signal-to-noise ratio is maximized at a cut value of > -0.05 where the signal efficiency is $\sim 77\%$ against a QCD jet mistag rate of $\sim 3\%$. The figure also shows it is possible to get a good separation against the background due to hadronic top jets, with the same discriminators.

Note that, a method to identify boosted top jets consisting of electrons was studied in [41]. The final outcome of our study is found to be comparable with the same as obtained in [41]. We thus see that our proposed top tagger has acceptable efficiency and can be used to study boosted objects consisting of b quark and τ lepton in the decays involving BSM particles in the context of BSM searches.

V. TAGGING A POLARIZED TOP

In the SM the t/\bar{t} quarks produced via QCD are essentially unpolarized due to the vector nature of QCD whereas for the single t production in association with a W , the $V-A$ nature of the tbW coupling completely determines the polarization of the produced t . Top quarks produced from BSM sources, either in pair or singly, may have a polarization different from the predictions of the SM depending on the chiral structure of the BSM vertices responsible for its production. Hence the polarization of the produced t is a good probe of many a BSM physics scenarios (see for example [2–5,37,45,67–72]). As already pointed out, in the rest frame of the top quark, the angular distribution of the decay products carries information about the initial top spin direction [16–18]. The angular distribution of the decay product (f) in the rest frame of the top with polarization \mathcal{P}_t , ($-1 \leq \mathcal{P}_t \leq 1$) is given by,

$$\frac{1}{\Gamma} \frac{d\Gamma}{d \cos \theta_f} = \frac{1}{2} (1 + k_f \mathcal{P}_t \cos \theta_f), \quad (5.1)$$

where k_f is the spin analyzing power and θ_f is the angle of the decay product f with respect to the top spin direction in the top rest frame. The down-type quark and charged lepton, originating from the decay of W -boson in hadronic and leptonic top decay, respectively, have the maximum power ($k_{\ell^+} = k_{\bar{d}} = 1$). This makes the leptonic decay mode of the t -quark particularly suited for t -polarization measurements. The decay product with the next highest analysing power is the b -quark and one has $\kappa_b = -\kappa_W = -0.41$. Note the opposite signs of κ_{ℓ^+} and κ_b .

In the boosted regime, the finite angular resolution of the detector reduces the effectiveness of the angular distributions of the decay products. Luckily one can construct polarization sensitive observables like energy fractions unambiguously without the requirement of W reconstruction inside the top jet [73,74] to study the polarization of top quark. To this end one can exploit the kinematic features of the top decay products [37,41,71,73–75]. Here in this section, we explore the feasibility of distinguishing between the left-, right- and un- polarized, boosted semileptonic top quarks using polarization sensitive observables. For that, we repeat the procedure of top tagging, i.e., steps 1–5 mentioned in the latter part of Sec. III and construct a few polarization sensitive variables constructed out of the energies and momenta of the tagged top jet and its subjects.

(i) Energy fractions:

The different angular distributions of the decay products in the rest frame of the t get translated into different energy distributions of the decay products in the lab-frame for the boosted t -quark and hence of course to different distributions in energy fractions of the decay t that these carry. The difference

in the distributions in these energy fractions (at the parton level this is just the ratio of the lepton or b quark energies to the energy of the t) for the left and right polarized top quarks was first pointed out in [73] and then used for tagging the polarized top jet in [74]. The left and right panels of Fig. 11 show the behavior of energy fraction variables (Z_j) defined in Eq. (3.2) for both the τ_h - and b - like jets respectively, for the case of left-handed top quarks (t_L), right-handed top quarks (t_R) and unpolarized top quarks (t_{LR}) originating from W' decay. The corresponding expected distributions at the parton level are shown by dotted lines for comparison. For a more realistic comparison with the jet level plots, one uses the visible energy from τ -lepton to determine Z_{τ_h} as described in Sec. III.

We do see the same difference, as seen in Fig. 6, between the energy fractions carried by the b -jets and τ_h -jets. Further, we note that b -like jets from t_L are more boosted compared to those from the t_R whereas for the τ_h , exactly opposite is the case. This can be of course understood in terms of the opposite signs of the spin analysing powers of the b quark and the τ^+ . Due to the negative sign of κ_b according to Eq. (5.1), b quarks are preferentially emitted opposite to the spin direction in the rest frame of the t . For the t_L , this means that they are emitted preferentially in the direction of motion of the t in the laboratory. This, in turn, means that the boost from the rest frame of the t to the laboratory frame makes these b quarks more energetic than would be the case with b quarks coming from the decay of t_R or unpolarized t -quarks. For the τ^+ the positive nature of κ_{τ} implies exactly the opposite. As can be

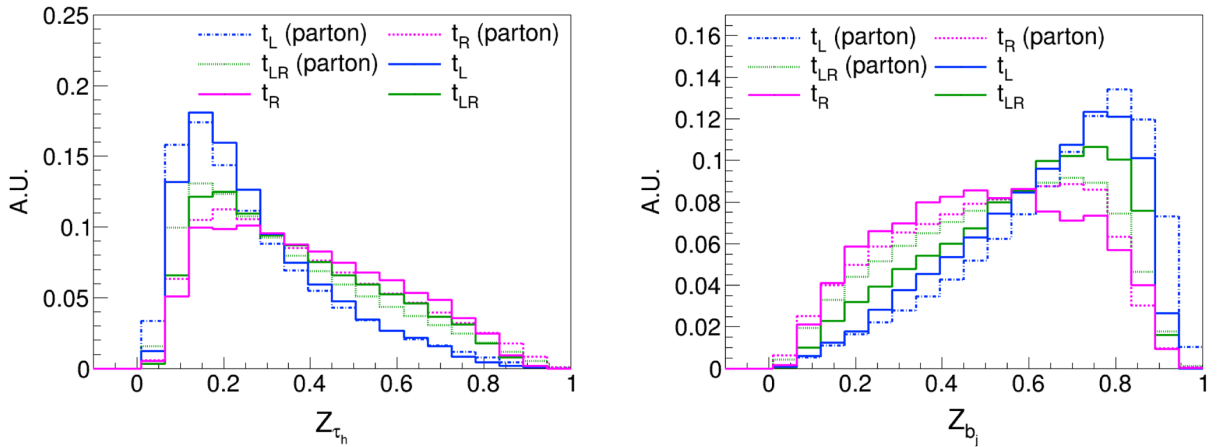


FIG. 11. Distribution of energy fraction [Eq. (3.2)] of τ_h - (left panel) and b - (right panel) like subjects for left-, right-, and unpolarized top. Blue solid (dotted) lines denote the distribution for left-handed reconstructed (parton level) top and magenta solid (dotted) is for right-handed reconstructed (parton level) top. Similarly, the green solid (dotted) distributions are for reconstructed (parton level) unpolarized top.

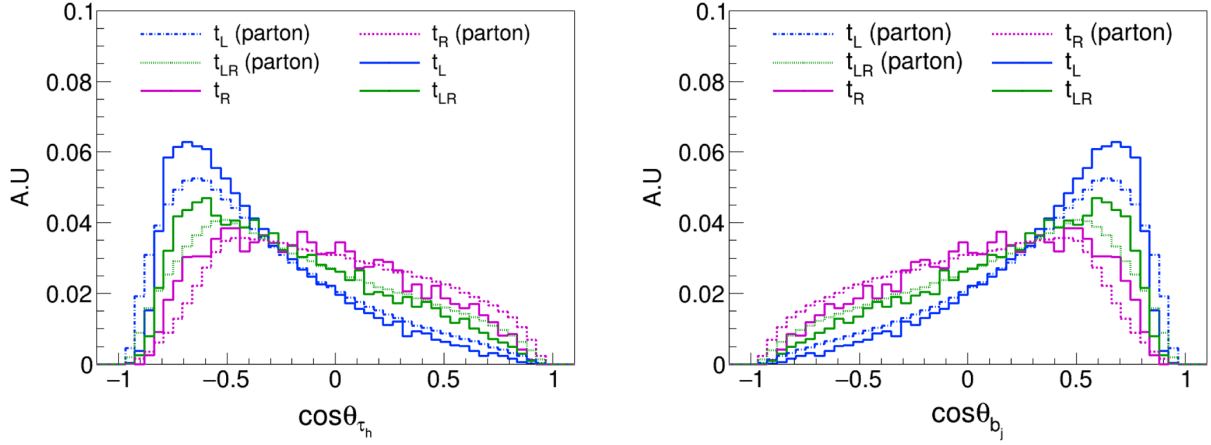


FIG. 12. The angular distribution of τ_h (left) and b_j (right) in the rest frame of $(b_j + \tau_h)$ system for t_L , t_R , and t_{LR} following Eq. (5.2). The color code is similar to Fig. 11.

seen from the left panel of Fig. 11, the τ -jet energy fraction (Z_{τ_h}) peaks ~ 0.15 for t_L , taking relatively larger values (~ 0.3 or more) for t_R . The right plot of Fig. 11, displays the characteristics of the b -jet that it takes away more energy for t_L than t_R . The distributions Z_{b_j} shown in the right panel of Fig. 11 peak at ~ 0.8 and 0.4 for t_L and t_R respectively. For comparison, we have included the distributions of energy fractions (Z_{τ_h} and Z_{b_j}) for t_{LR} which we have previously discussed in Fig. 6. Thus, we see that even though the full reconstruction of the τ momentum is not possible at the jet level (due to the presence of neutrinos in the final state), the energy fractions can act as a good polarimeter for differentiating left-handed, right-handed and unpolarized boosted top quarks decaying to final state containing τ leptons.

(ii) Angular variable ($\cos\theta_j$):

Equation (5.1) gives the distribution in the angle of the decay product with the spin direction of the decaying t , in the rest frame of the t . Since for the t -quark of a given helicity, the spin direction is related to the direction of the t three-momentum vector; we can instead look at the correlation of the decay product momentum in the rest frame of the t , with the original t direction in the laboratory. In case of the hadronic topjet, the topjet direction in the laboratory is of course a very good proxy for the momentum of the t -quark in laboratory. This was used to good effect in [71] to discriminate boosted topjets with different helicities. In the present case due to the presence of missing momentum, we consider sum of the reconstructed momentum of the $b - \tau$ jet system in the laboratory frame, viz. $(\vec{b}_j + \vec{\tau}_h)$ system as a proxy for the t momentum (and hence top-spin) direction. We then define

$$\cos\theta_j|_{J=b_j,\tau_h} = \frac{(\vec{b}_j + \vec{\tau}_h) \cdot \vec{j}'}{|\vec{b}_j + \vec{\tau}_h| |\vec{j}'|}. \quad (5.2)$$

Here j' is the momentum of the subjet (either b -jet or τ -jet) in the rest frame of the $b - \tau$ jet system.

For left-handed top quark, the direction of the top quark spin is in the opposite direction with respect to the top quark momentum in the laboratory frame. Equation (5.1) and the values of κ_{ℓ^+} , κ_b , tell us then that the $\tau_h(b)$ from t_R would be preferentially emitted along (opposite) the direction of the $(\vec{b}_j + \vec{\tau}_h)$, i.e., the proxy t momentum direction. Exactly opposite will be the case for t_L , and for t_{LR} it will be in between t_L and t_R . This is borne out by the left plot in Fig. 12, which shows that the $\cos\theta_{\tau_h}$ is preferentially positive (negative) for right (left) handed top quarks. As seen in the right plot in Fig. 12, the b -jet exhibits a behavior exactly opposite to the τ -case i.e., $\cos\theta_{b_j}$ takes preferentially negative (positive) values for right (left) handed top quark.

Using only these two sets of variables, namely the energy fractions (Z_{b_j} and Z_{τ_h}) and angular correlation discriminator ($\cos\theta_j$) of the top decay products, we can achieve a good separation between the left-handed, right-handed and unpolarized top quarks. Note that the top tagging efficiencies for left and right polarized top quark do differ only slightly $\sim 2\%$ – 3% . In Fig. 13, we display the ROC curve obtained by training the BDT with two polarization sensitive observables along with the other variables of Table III, for left, right and unpolarized top samples. The blue (magenta) curve shows the efficiency of t_L (t_R) against t_{LR} , while the yellow line represents the same for t_R against t_L . It can be seen from Fig. 13 that t_R is identified more efficiently than t_L against an unpolarized top background. This can be understood by looking at kinematic regions

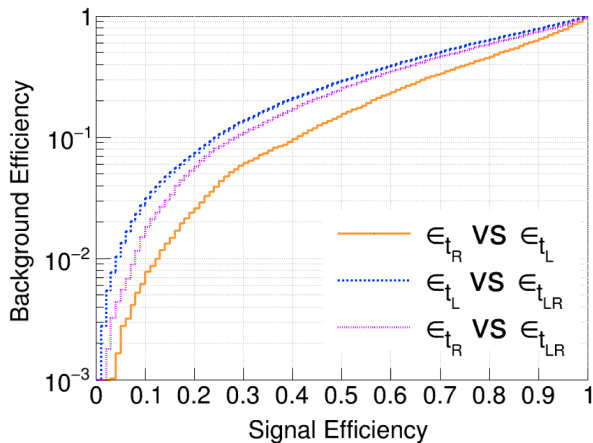


FIG. 13. The ROC curve estimates the performance of the BDT classifier of distinguishing the (a) right vs left (b) left vs unpolarized (c) right vs unpolarized top jets.

where one of the t_L or t_R differs distinctly from t_{LR} . In this case, the b -jet energy fraction seems to be a better discriminator between the polarized top and unpolarized sample for the right polarized top. As we pointed out previously, because a fraction of the energy of top is carried out by the MET in τ decay, Z_{b_j} peaks at higher values than Z_{τ_h} for t_{LR} . Hence, t_R for which the b quark from t decay is less energetic than that of t_L , has a higher chance of discrimination against t_{LR} . We find that right-handed top jets can be tagged with $\sim 65\%$ efficiency (ϵ_{t_R}) with a mistagging rate of $\sim 25\%$ for left-handed top jets (ϵ_{t_L}) and $\sim 35\%$ against unpolarized top ($\epsilon_{t_{LR}}$) jets. If we reduce ϵ_{t_L} to $\sim 1\%$ – 2% , ϵ_{t_R} comes out to be around 15% – 20% .

VI. SUMMARY

In this paper, we investigated the performance of a boosted top tagger when the top quark decays in the leptonic channel with the τ lepton in the final state. The proposed top tagger relies on the identification of an energetic b -jet and a τ -jet within the top large- R jet and its energy profile with respect to the parent particle. This methodology focuses on the distribution of energy between two subjets within the large- R jet and constructs kinematic variables relevant to the final state topology. This in turn helps to discriminate the signal from SM backgrounds. Some of the variables that are found to be useful in identifying and classifying the semileptonic boosted tops are (a) the energy fraction of the identified b - and τ - subjets of the total large- R jet energy, (b) the difference of the masses of the b – τ system and top large- R jet [cf. Eq. (3.3)], (c) the ratios of the N-subjettiness variables, and (d) transverse mass of the b – τ subject.

We analyze the performance of the proposed tagger by using simulated signal and background events and then constructing observables based on the jet substructure technique. Through a BDT analysis, we obtain a signal efficiency of around 77% while keeping the mistagging rate of the QCD jets (consists of quark and gluon initiated jets) to 3% level. The hadronically decaying top quark initiated jets can also play the role of a potential background. We find that with a signal efficiency of around 77% , the mistagging rate of the hadronic top quark jet is around 5% – 6% . It is noteworthy that even though the main focus of this study is to develop a top tagger when top quark decays through τ -lepton in the final state, however, this tagger can be applied to any large- R jet, which includes a b -jet and a τ -jet. For example, decays of a 3rd generation Leptoquark, top squark decay in R-parity violating supersymmetric models etc. can all lead to a final containing a b - and a τ - subjet inside a large- R jet due to the boosted nature of the parent particle. Absence of a ν in the primary decay in the signal ought to make the tagging using this strategy even easier.

Another important aspect of developing a top tagger for this leptonic final state, is that we can analyze and estimate the sensitivity of the tagger with respect to top quark polarization. The couplings of the top quark with fermions and bosons in various BSM scenarios can have different implications for the top quark polarization. This polarization in turn modifies the distribution of various kinematic observables involving its decay products. We study the sensitivity of these distributions to t -polarization in two extreme cases of top polarization, namely purely left-handed and purely right-handed top quarks. We use the same method of top-tagging as described above. We indeed find that using the observables based on the energy profile of subjets of the boosted top jet, namely the b - and τ - tagged jets along with the angular correlations measured in the rest frame of b – τ system, one can differentiate between the left and right-handed top quarks quite efficiently. A detailed exploration of some of the interesting applications of our proposed top tagger is left for future investigations.

ACKNOWLEDGMENTS

The work of A. C. is funded by the Department of Science and Technology, Government of India, under Grant No. IFA18-PH 224 of Innovation in Science Pursuit for Inspired Research (INSPIRE) Faculty Award. R. M. G. wishes to acknowledge the support of Indian National Science Academy under the award of INSA Senior Scientist. A. D. thanks Ritesh Kumar Singh (IISER, Kolkata) for helpful discussions. A. D. would like to acknowledge the support of DST-INSPIRE Fellowship (IF160414).

- [1] R. Kehoe, M. Narain, and A. Kumar, Review of top quark physics results, *Int. J. Mod. Phys. A* **23**, 353 (2008).
- [2] P. S. Bhupal Dev, A. Djouadi, R. M. Godbole, M. M. Muhlleitner, and S. D. Rindani, Determining the CP Properties of the Higgs Boson, *Phys. Rev. Lett.* **100**, 051801 (2008).
- [3] Saurabh D. Rindani and Pankaj Sharma, Probing anomalous tbW couplings in single-top production using top polarization at the large hadron collider, *J. High Energy Phys.* **11** (2011) 082.
- [4] Saurabh D. Rindani, Rui Santos, and Pankaj Sharma, Measuring the charged Higgs mass and distinguishing between models with top-quark observables, *J. High Energy Phys.* **11** (2013) 188.
- [5] Fawzi Boudjema, Rohini M. Godbole, Diego Guadagnoli, and Kirtimaan A. Mohan, Lab-frame observables for probing the top-Higgs interaction, *Phys. Rev. D* **92**, 015019 (2015).
- [6] Hans Peter Nilles, Supersymmetry, supergravity and particle physics, *Phys. Rep.* **110**, 1 (1984).
- [7] J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, Princeton, NJ, USA, 1992).
- [8] Howard E. Haber and Gordon L. Kane, The search for supersymmetry: Probing physics beyond the standard model, *Phys. Rep.* **117**, 75 (1985).
- [9] M. Drees, R. Godbole, and P. Roy, *Theory and Phenomenology of Sparticles: An Account of Four-Dimensional $N = 1$ Supersymmetry in High Energy Physics* (World Scientific, Singapore, 2004), <https://doi.org/10.1142/4001>.
- [10] H. Baer and X. Tata, *Weak Scale Supersymmetry: From Superfields to Scattering Events* (Cambridge University Press, Cambridge, England, 2006).
- [11] David B. Kaplan and Howard Georgi, $SU(2) \times U(1)$ breaking by vacuum misalignment, *Phys. Lett.* **136B**, 183 (1984).
- [12] Martin Schmaltz and David Tucker-Smith, Little Higgs review, *Annu. Rev. Nucl. Part. Sci.* **55**, 229 (2005).
- [13] Lisa Randall and Raman Sundrum, Large Mass Hierarchy from a Small Extra Dimension, *Phys. Rev. Lett.* **83**, 3370 (1999).
- [14] Ulrich Husemann, Top-quark physics: Status and prospects, *Prog. Part. Nucl. Phys.* **95**, 48 (2017).
- [15] M. Beneke *et al.*, Top quark physics, in *Proceedings of the Workshop on Standard Model Physics (and More) at the LHC (First Plenary Meeting)* (2000), pp. 419–529, [10.5170/CERN-2000-004.419](https://arxiv.org/abs/10.5170/CERN-2000-004.419).
- [16] M. Jezabek and Johann H. Kuhn, Lepton spectra from heavy quark decay, *Nucl. Phys.* **B320**, 20 (1989).
- [17] M. Jezabek, Top quark physics, *Nucl. Phys. B, Proc. Suppl.* **37**, 197 (1994).
- [18] Werner Bernreuther, Top quark physics at the LHC, *J. Phys. G* **35**, 083001 (2008).
- [19] Saurabh D. Rindani, Effect of anomalous tbW vertex on decay lepton distributions in $e^+e^- \rightarrow tt$ and CP violating asymmetries, *Pramana* **54**, 791 (2000).
- [20] Bohdan Grzadkowski and Zenro Hioki, New hints for testing anomalous top quark interactions at future linear colliders, *Phys. Lett. B* **476**, 87 (2000).
- [21] Zenro Hioki, Probing anomalous top couplings through the final lepton angular and energy distributions at polarized NLC, in *Proceedings of the Theory Workshop on Physics at Linear Colliders, Tsukuba, Japan (C01-03-15)* (2001), pp. 169–177, [arXiv:hep-ph/0104105](https://arxiv.org/abs/hep-ph/0104105).
- [22] Kazumasa Ohkuma, Effects of top quark anomalous decay couplings at $\gamma\gamma$ colliders, *Nucl. Phys. B, Proc. Suppl.* **111**, 285 (2002).
- [23] Bohdan Grzadkowski and Zenro Hioki, Angular distribution of leptons in general $t\bar{t}$ production and decay, *Phys. Lett. B* **529**, 82 (2002).
- [24] Bohdan Grzadkowski and Zenro Hioki, Decoupling of anomalous top decay vertices in angular distribution of secondary particles, *Phys. Lett. B* **557**, 55 (2003).
- [25] Zenro Hioki, A new decoupling theorem in top quark physics, in *Proceedings of the International Workshop on Linear Colliders (LCWS 2002), Jeju Island, Republic of Korea (C02-08-26.5)* (2002), pp. 333–338, [arXiv:hep-ph/0210224](https://arxiv.org/abs/hep-ph/0210224).
- [26] Rohini M. Godbole, Saurabh D. Rindani, and Ritesh K. Singh, Study of CP property of the Higgs at a photon collider using $\gamma\gamma \rightarrow t\bar{t} \rightarrow lX$, *Phys. Rev. D* **67**, 095009 (2003); *Phys. Rev. D* **71**, 039902(E) (2005).
- [27] Rohini M. Godbole, Saurabh D. Rindani, and Ritesh K. Singh, Lepton distribution as a probe of new physics in production and decay of the t quark and its polarization, *J. High Energy Phys.* **12** (2006) 021.
- [28] Rohini M. Godbole, Kumar Rao, Saurabh D. Rindani, and Ritesh K. Singh, On measurement of top polarization as a probe of $t\bar{t}$ production mechanisms at the LHC, *J. High Energy Phys.* **11** (2010) 144.
- [29] Rohini M. Godbole, Michael E. Peskin, Saurabh D. Rindani, and Ritesh K. Singh, Why the angular distribution of the top decay lepton is unchanged by anomalous tbW couplings, *Phys. Lett. B* **790**, 322 (2019).
- [30] Gavin P. Salam, Towards jetography, *Eur. Phys. J. C* **67**, 637 (2010).
- [31] Jesse Thaler and Lian-Tao Wang, Strategies to identify boosted tops, *J. High Energy Phys.* **07** (2008) 092.
- [32] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie, Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks, *Phys. Rev. Lett.* **101**, 142001 (2008).
- [33] Tilman Plehn and Michael Spannowsky, Top tagging, *J. Phys. G* **39**, 083001 (2012).
- [34] Leandro G. Almeida, Mihailo Backović, Mathieu Cliche, Seung J. Lee, and Maxim Perelstein, Playing tag with ANN: Boosted top identification with pattern recognition, *J. High Energy Phys.* **07** (2015) 086.
- [35] Gregor Kasieczka, Tilman Plehn, Anja Butter, Kyle Cranmer, Dipsikha Debnath, Barry M. Dillon, Malcolm Fairbairn, Darius A. Faroughy, Wojtek Fedorko, Christophe Gay *et al.*, The machine learning landscape of top taggers, *SciPost Phys.* **7**, 014 (2019).
- [36] Gregor Kasieczka, Tilman Plehn, Michael Russell, and Torben Schell, Deep-learning top taggers or the end of QCD?, *J. High Energy Phys.* **05** (2017) 006.
- [37] Soham Bhattacharya, Monoranjan Guchait, and Aravind H. Vijay, Boosted top quark tagging and

- polarization measurement using machine learning, *Phys. Rev. D* **105**, 042005 (2022).
- [38] Biplob Bhattacharjee, Camellia Bose, Amit Chakraborty, and Rhitaja Sengupta, Boosted top tagging and its interpretation using Shapley values, [arXiv:2212.11606](https://arxiv.org/abs/2212.11606).
- [39] Andrew J. Larkoski, Ian Moulton, and Benjamin Nachman, Jet substructure at the large hadron collider: A review of recent advances in theory and machine learning, *Phys. Rep.* **841**, 1 (2020).
- [40] Roman Kogler *et al.*, Jet substructure at the large hadron collider: Experimental review, *Rev. Mod. Phys.* **91**, 045003 (2019).
- [41] Suman Chatterjee, Rohini Godbole, and Tuhin S. Roy, Jets with electrons from boosted top quarks, *J. High Energy Phys.* **01** (2020) 170.
- [42] B. K. Bullock, Kaoru Hagiwara, and Alan D. Martin, Tau polarization and its correlations as a probe of new physics, *Nucl. Phys.* **B395**, 499 (1993).
- [43] R. M. Godbole, Monoranjan Guchait, and D. P. Roy, Using tau polarization to probe the stau co-annihilation region of the minimal supergravity model at the LHC, *Phys. Rev. D* **79**, 095015 (2009).
- [44] Monoranjan Guchait and D. P. Roy, Using tau polarisation for charged Higgs boson and SUSY searches at the LHC, in *Physics at the Large Hadron Collider* (Springer, New Delhi, 2008), pp. 205–212, [10.1007/978-81-8489-295-6_13](https://doi.org/10.1007/978-81-8489-295-6_13).
- [45] Thomas Gajdosik, Rohini M. Godbole, and Sabine Kraml, Fermion polarization in sfermion decays as a probe of CP phases in the MSSM, *J. High Energy Phys.* **09** (2004) 051.
- [46] Lawrence J. Hall and Mahiko Suzuki, Explicit R-parity breaking in supersymmetric models, *Nucl. Phys.* **B231**, 419 (1984).
- [47] J. C. Romao, M. A. Diaz, M. Hirsch, W. Porod, and J. W. F. Valle, A supersymmetric solution to the solar and atmospheric neutrino problems, *Phys. Rev. D* **61**, 071703 (2000).
- [48] Eung Jin Chun, Sunghoon Jung, Hyun Min Lee, and Seong Chan Park, Stop and sbottom LSP with R-parity violation, *Phys. Rev. D* **90**, 115023 (2014).
- [49] Zack Sullivan, Fully differential W' production and decay at next-to-leading order in QCD, *Phys. Rev. D* **66**, 075011 (2002).
- [50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *J. High Energy Phys.* **07** (2014) 079.
- [51] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk, Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, *J. High Energy Phys.* **03** (2013) 015.
- [52] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, An introduction to PYTHIA 8.2, *Comput. Phys. Commun.* **191**, 159 (2015).
- [53] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [54] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, The anti- k_r jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, Fastjet user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [56] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler, Soft drop, *J. High Energy Phys.* **05** (2014) 146.
- [57] Albert M. Sirunyan *et al.*, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, *J. Instrum.* **15**, P06005 (2020).
- [58] Giuseppe Bagliesi, Tau tagging at ATLAS and CMS, in *Proceedings of the 17th Symposium on Hadron Collider Physics 2006 (HCP 2006), Durham, NC, USA (C06-05-22.1)* (2007), [arXiv:0707.0928](https://arxiv.org/abs/0707.0928).
- [59] Andrey Katz, Minh Son, and Brock Tweedie, Ditaup tagging and boosted Higgs from a multi-TeV resonance, *Phys. Rev. D* **83**, 114033 (2011).
- [60] S. Gennai, F. Moortgat, L. Wendland, A. Nikitenko, S. Wakefield, G. Bagliesi, S. Dutta, A. Kalinowski, M. Konecki, and D. Kotlinski, Tau jet reconstruction and tagging at high level trigger and off-line, *Eur. Phys. J. C* **46**, 1 (2006).
- [61] The CMS Collaboration, Tau jet reconstruction and tagging with CMS, *Eur. Phys. J. C* **46**, 1 (2006).
- [62] Georges Aad *et al.*, ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79**, 970 (2019).
- [63] CMS collaboration, Identification of b-quark jets with the CMS experiment, *J. Instrum.* **8**, P04013 (2013).
- [64] Jesse Thaler and Ken Van Tilburg, Identifying boosted objects with n-subjettiness, *J. High Energy Phys.* **03** (2011) 015.
- [65] Andreas Hocker *et al.*, TMVA—Toolkit for multivariate data analysis, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [66] Yoav Freund and Robert E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* **55**, 119 (1997).
- [67] G. Belanger, R. M. Godbole, L. Hartgring, and I. Niessen, Top polarization in stop production at the LHC, *J. High Energy Phys.* **05** (2013) 167.
- [68] Rohini M. Godbole, Lisa Hartgring, Irene Niessen, and Chris D. White, Top polarisation studies in H^-t and Wt production, *J. High Energy Phys.* **01** (2012) 011.
- [69] Rohini M. Godbole, Gaurav Mendiratta, and Saurabh Rindani, Looking for bSM physics using top-quark polarization and decay-lepton kinematic asymmetries, *Phys. Rev. D* **92**, 094013 (2015).
- [70] V. Arunprasath, Rohini M. Godbole, and Ritesh K. Singh, Polarization of a top quark produced in the decay of a gluino or a stop in an arbitrary frame, *Phys. Rev. D* **95**, 076012 (2017).
- [71] Rohini Godbole, Monoranjan Guchait, Charanjit K. Khosa, Jayita Lahiri, Seema Sharma, and Aravind H. Vijay, Boosted top quark polarization, *Phys. Rev. D* **100**, 056010 (2019).
- [72] Genevieve Belanger, Rohini M. Godbole, Charanjit K. Khosa, and Saurabh D. Rindani, Probing CP nature of a

- mediator in associated production of dark matter with single top quark, in *Proceedings of the 11th International Workshop on Top Quark Physics (TOP2018), Bad Neuenahr, Germany (C18-09-16.1)* (2018), [arXiv:1811.11048](https://arxiv.org/abs/1811.11048).
- [73] Jessie Shelton, Polarized tops from new physics: Signals and observables, *Phys. Rev. D* **79**, 014032 (2009).
- [74] David Krohn, Jessie Shelton, and Lian-Tao Wang, Measuring the polarization of boosted hadronic tops, *J. High Energy Phys.* **07** (2010) 041.
- [75] Yoshio Kitadono and Hsiang-nan Li, Jet substructures of boosted polarized hadronic top quarks, *Phys. Rev. D* **93**, 054043 (2016).