

Improving sensitivity of vectorlike top partner searches with jet substructure

Anupam Ghosh^{1,*}, Soumyadip Ghosh^{2,†}, Soureek Mitra^{3,‡}, Tousik Samui^{4,§} and Ritesh K. Singh^{2,||}

¹*Department of Physics, Indian Institute of Technology Guwahati, North Guwahati, 781039 Assam, India*

²*Department of Physical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, 741246 Nadia, India*

³*Department of Physical Sciences, Indian Institute of Science Education and Research, Knowledge City, Sector 81, S. A. S. Nagar, Manauli PO 140306 Punjab, India*

⁴*The Institute of Mathematical Sciences, IV Cross Road, CIT Campus, Taramani, Chennai 600113, India*



(Received 24 October 2025; accepted 23 January 2026; published 26 February 2026)

Vectorlike quark partners appear in many models beyond the Standard Model and remain an important area of research, as they can offer insights into the electroweak symmetry breaking mechanism. In this work, we have focused on studying the production of a heavy vectorlike top partner in association with a Standard Model top quark via chromomagnetic coupling and the four decay modes of the top partner, namely, bW , tZ , th , and tg . The signal has been studied in final states with one fat jet, at least one b -tagged jet, one lepton, and missing energy. This study focuses on the extensive use of jet substructure techniques in jets clustered with fixed and dynamically varying radius to deal with events containing differently sized jets. Important kinematic information, along with jet substructure and event shape observables, has been used in a multivariate analysis to extract signals with high significance. A comparative study between fixed and dynamically varying radius clustering of jets is also presented, leading to an improvement in signal sensitivity in the highly boosted scenario.

DOI: [10.1103/dhs6-gcn9](https://doi.org/10.1103/dhs6-gcn9)

I. INTRODUCTION

The search for new physics beyond the Standard Model (BSM) remains one of the most important goals of current and future high-energy collider experiments. Despite many successes, the Standard Model (SM) of elementary particles leaves several fundamental questions unanswered, such as the nature of electroweak symmetry breaking and the stability of the electroweak vacuum [1–5], the origin of mass hierarchies [6,7], matter-antimatter asymmetry [8,9], the composition of dark matter [10,11], etc. Many BSM models [12–14] propose the existence of new resonances, such as vectorlike fermions [15–18], that can provide critical insights into some of these open problems and offer experimentally verifiable predictions at the TeV scale [19]. Vectorlike fermions are directly related to the

stability of the electroweak vacuum [20]. A vectorlike top partner, being a color triplet and having the same electric charge as the SM top quark, couples to the Higgs boson and alters the renormalization group evolution of the Higgs quartic coupling. Its Yukawa interaction contributes positively to the running of the Higgs quartic coupling, while its mixing with the SM top quark reduces the Yukawa contribution due to the top quark that typically drives the coupling negative. Together, these effects slow down or prevent the Higgs quartic coupling from turning negative at high-energy scales, thus stabilizing the electroweak vacuum all the way up to the Planck scale [20]. In addition, colored vectorlike fermions can mediate interactions between dark matter and SM quarks as indicated in the phenomenological analyses of vectorlike portal dark matter models [21–25]. Thus, identifying such BSM states would not only signify the discovery of new physics phenomena but also provide a profound understanding of the structure of fundamental interactions.

Extraction of rare BSM signals from overwhelming SM backgrounds requires improved analysis techniques and theoretical modeling. State-of-the-art jet substructure methods [26–28], clustering algorithms [29,30], and machine-learning techniques [31,32] have become indispensable tools for identifying boosted heavy resonances and complex event topologies that are characteristic of various BSM

*Contact author: anupamg@rnd.iitg.ac.in

†Contact author: sg22rs009@iiserkol.ac.in

‡Contact author: soureekmitra@iisermohali.ac.in

§Contact author: tousiks@imsc.res.in

||Contact author: ritesh.singh@iiserkol.ac.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.

signatures at the Large Hadron Collider (LHC). At the same time, access to high-quality data and simulation samples, such as CMS Open Data [33,34], allows for a more reliable background model and realistic performance estimates.

The nature of modern BSM searches requires exploration at increasingly higher-energy regimes. At these high energies, fat jets frequently arise from the hadronic decays of boosted heavy particles such as the top quarks or the heavy bosons, after parton showering and hadronization. One of the primary aspects of this paper is to assess the efficacy of the recently proposed “dynamic-radius” (DR) jet clustering algorithm [35], which is advantageous in the higher-energy regime where such fat jets are prominent. The substructure of the fat jets originating from massive boosted particle decays differs significantly from that of narrow QCD jets [36] owing to their decay history. The DR algorithm leverages this substructure—more precisely, combinations of energy correlation functions [37] to adaptively set the jet radius based on local kinematics, thus producing variable-sized jets in an event. Some recent studies have made use of the DR clustering algorithm with a zero pileup scenario [38–40]. However, understanding the performance of this algorithm in realistic experimental conditions remains extremely crucial.

In this work, we investigate the production of a heavy vectorlike top partner (T) in association with a SM top quark, mediated through chromomagnetic interactions [41–43]. The top partner decays through four main channels: tZ , th , tg , and bW ; each leading to distinct final-state signatures.¹ Among these, direct searches at the LHC have been performed primarily in the tZ and th channels, leading to stringent lower bounds on the mass of the top partner around a few TeV [44–46]. Furthermore, indirect constraints on the BSM nature of the tg coupling have been derived using $t\bar{t}$ events [47,48]. However, only a handful of direct searches in the bW decay mode have been performed at the LHC with full Run-II data [49,50] due to its limited sensitivity. The final state for this study consists of exactly one lepton, exactly one fat jet, at least one narrow jet that is b tagged, and missing transverse momentum (p_T^{miss}), capturing event topologies sensitive to the presence of vectorlike top partners. The bW decay mode of the T quark constitutes a significant contribution to this specific final state. The probe for the signal process is conducted using jets clustered using traditional fixed radius as well as the more recent DR approaches. We employ powerful jet substructure tools, namely, soft drop (SD) grooming [51], N subjettiness [52], and energy correlation functions (ECFs) [37] to better understand the nature and the behavior of the fat jet in the event. In addition, we also studied several event shape

variables, such as sphericity [53,54], aplanarity [55], Fox-Wolfram moments (H) [56,57], etc., along with lepton and jet kinematic information in a multivariate analysis, to effectively distinguish signal events from the background. The signal events are simulated with MadGraph5_aMC@NLO [58,59] together with Pythia8 [60] for hadronization and parton showering and Delphes [61] for detector response modeling. For SM backgrounds, high-statistics Monte Carlo (MC) samples from CMS Open Data have been utilized for a reliable estimate. This study also includes pileup simulation with a realistic value of the mean pileup interactions, considering the Run-II LHC conditions.

Before proceeding further, it should be noted that the main goal of this work is to explore the usefulness of the DR jet clustering in highly boosted final states and to compare its performance with standard fixed-radius clustering. Impact of various uncertainty sources, such as pileup contamination removal, jet energy scale and resolution, hadronization model, etc., on the DR algorithm requires dedicated in-depth studies that are beyond the scope of the present work. This study is intended as a proof of concept, focusing on qualitative improvements achievable with DR jet clustering using a purely statistical approach. This is an initial step to demonstrate the applicability of the DR jet clustering, considering a well-motivated candidate BSM process involving a heavy top partner coupled strongly with the SM top quark via chromomagnetic coupling.

The article is organized as follows. Section II A introduces the signal model based on chromomagnetic coupling between the vectorlike top partner and the SM top quark, while the description of signal and background simulation procedures is presented in Secs. II B and II C, respectively. The DR algorithm and relevant jet substructure techniques are presented in Sec. III together with a control study using $t\bar{t} + \text{jets}$ events. Section IV discusses the event reconstruction and selection strategy. In Sec. V, we detail the multivariate analysis strategy, optimization methods, and present our main results. Finally, Sec. VI provides a summary of the work and an outlook toward the future prospects.

II. SIGNAL AND BACKGROUND PROCESSES

A. Signal model

The vectorlike T quark is a triplet under $SU(3)$, and its electric charge is determined by its hypercharge and isospin structure. Additionally, it can couple with any generation of the SM quark. We assume the heavy top partner predominantly interacts with the third-generation SM quarks. The T quark is a singlet under weak isospin and has the same electromagnetic charge as the top quark; hence, it mixes with the top quark. The general Lagrangian terms are given as follows [41–43]:

¹In this article, we followed a convention to denote all SM particles in roman and BSM particles in italics. However, the associated fields of both SM and BSM particles are denoted in italics.

$$\mathcal{L}_{TWb} = -\frac{g_w \sin \theta_L}{\sqrt{2}} \bar{T} \gamma^\mu P_L b W_\mu + \text{H.c.}, \quad (1)$$

$$\mathcal{L}_{TZt} = \frac{g_w}{2c_w} \bar{t} \gamma^\mu (2 \sin \theta_L \cos \theta_L P_L + \sin \theta_R \cos \theta_R P_R) T Z_\mu + \text{H.c.}, \quad (2)$$

$$\mathcal{L}_{Th} = -s_w \frac{\sin \theta_R \cos \theta_R}{2M_W} \bar{t} (M_T P_L + m_t P_R) T h + \text{H.c.} \quad (3)$$

Here, t and h denote the SM top quark and Higgs boson, respectively. $P_{L/R} = \frac{1 \mp \gamma_5}{2}$ represent the left- and right-chiral projection operators. The mixing angle between the left-chiral (right-chiral) SM top quark and the left-chiral (right-chiral) T quark is denoted by θ_L (θ_R). The symbols M_T , m_t , and M_W represent the masses of the T quark, the SM top quark, and the W boson, respectively. The $SU(2)_L$ gauge coupling is g_w ; s_w and c_w are sine and cosine of the electroweak mixing angle. The single production of the exotic T quark in association with the SM top quark is possible at the LHC via gluon fusion. The effective Lagrangian of such interactions is as follows:

$$\mathcal{L}_{tg} = \frac{g_s}{\Lambda} G_{\mu\nu}^a [C_{tRL} \bar{T} \sigma^{\mu\nu} \tau^a P_L t + C_{tLR} \bar{T} \sigma^{\mu\nu} \tau^a P_R t] + \text{H.c.}, \quad (4)$$

where $\sigma_{\mu\nu} = i/2(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu)$, and g_s is the SM strong coupling constant. The notation $G_{\mu\nu}^a$ represents the gluon field strength tensor, and τ^a corresponds to the $SU(3)$ generators. The notation Λ denotes the new physics scale, and C_{tRL} , C_{tLR} are the coupling strengths. Corresponding to the above Lagrangian given in Eqs. (1)–(4), the T quark has four decay channels, namely, bW , tZ , th , and tg . While Eq. (4) primarily contributes to the production of the T quark, it is also responsible for the tg decay mode. From Eqs. (1)–(3), one notices that the branching ratios for $T \rightarrow tZ$, $T \rightarrow bW$, and $T \rightarrow th$ become negligibly small when the mixing angles are very small ($\theta_L, \theta_R \rightarrow 0$), and the T quark decays predominantly into a top quark and a gluon. However, with increasing mixing between the T quark and the top quark, the branching fractions into the electroweak channels $T \rightarrow tZ$, $T \rightarrow bW$, and $T \rightarrow th$ increase significantly.

B. Signal simulation

The process we are interested in is the associated production of a SM top quark and its vectorlike partner T via the effective chromomagnetic interaction,

$$pp \rightarrow t\bar{T} + X \quad \text{and} \quad (5)$$

$$pp \rightarrow T\bar{t} + X. \quad (6)$$

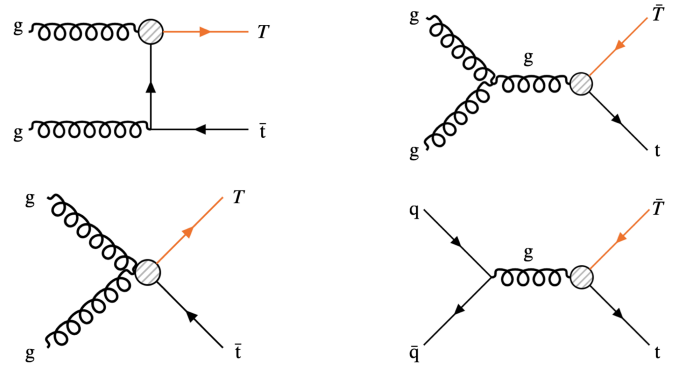


FIG. 1. Representative parton-level Feynman diagrams for the process $pp \rightarrow T\bar{t}/\bar{T}t$.

A set of representative parton-level Feynman diagrams corresponding to the process has been given in Fig. 1.

To simulate the processes indicated by Eqs. (5) and (6), we first implemented all the interaction terms defined in Eqs. (1)–(4), along with the SM Lagrangian, using FeynRules [62,63] to generate a Universal FeynRules Output (UFO) [64] model file. This UFO model was then used in the MadGraph5_aMC@NLO framework to generate parton-level events at a center-of-mass energy of $\sqrt{s} = 13$ TeV with the parton distribution function (PDF) set to NNPDF30_nlo_nf_5_pdfas PDFset [65]. The factorization and renormalization scales are set to the default dynamical scale in MadGraph5, which corresponds to the transverse mass of the $2 \rightarrow 2$ scattering process at leading order (LO). Figure 2 shows the variation of the partonic cross section for T quark production in association with a top quark at $\sqrt{s} = 13$ TeV, along with the branching ratios of its various decay channels, for a fixed new physics scale of $\Lambda = 8$ TeV.

To illustrate the collider analysis for this study, we consider a benchmark point that is consistent with existing experimental bounds. The strongest lower bound on M_T is about 1.8 TeV [44]. We further consider $\theta_L = \theta_R$ and $C_{tRL} = C_{tLR}$ to reduce the number of free parameters. The couplings C_{tRL} and C_{tLR} control the top-partner pair production cross section, which is restricted to $\mathcal{O}(1 \text{ fb})$ [19], resulting in an upper limit on $\frac{C_{tRL}}{\Lambda} = \frac{C_{tLR}}{\Lambda} \lesssim 0.6 \text{ TeV}^{-1}$. Furthermore, the mixing parameters θ_L and θ_R affect the Higgs boson branching ratio to gluon-gluon and diphoton modes. One finds $\theta_L = \theta_R \lesssim 0.25$ rad for $M_T = 2.2$ TeV from Refs. [17,66,67]. Based on these factors, we choose the benchmark scenario as follows:

$$\frac{C_{tRL}}{\Lambda} = \frac{C_{tLR}}{\Lambda} = 0.1 \text{ TeV}^{-1}, \quad \theta_L = \theta_R = 0.06 \text{ rad}, \quad \text{and } M_T = 2.2 \text{ TeV}. \quad (7)$$

The parton-level kinematics of the produced top partner are shown in Fig. 3 for different mass hypotheses. The $p_T(T)$ distribution peaks around $p_T(T) \sim 200$ GeV, with a

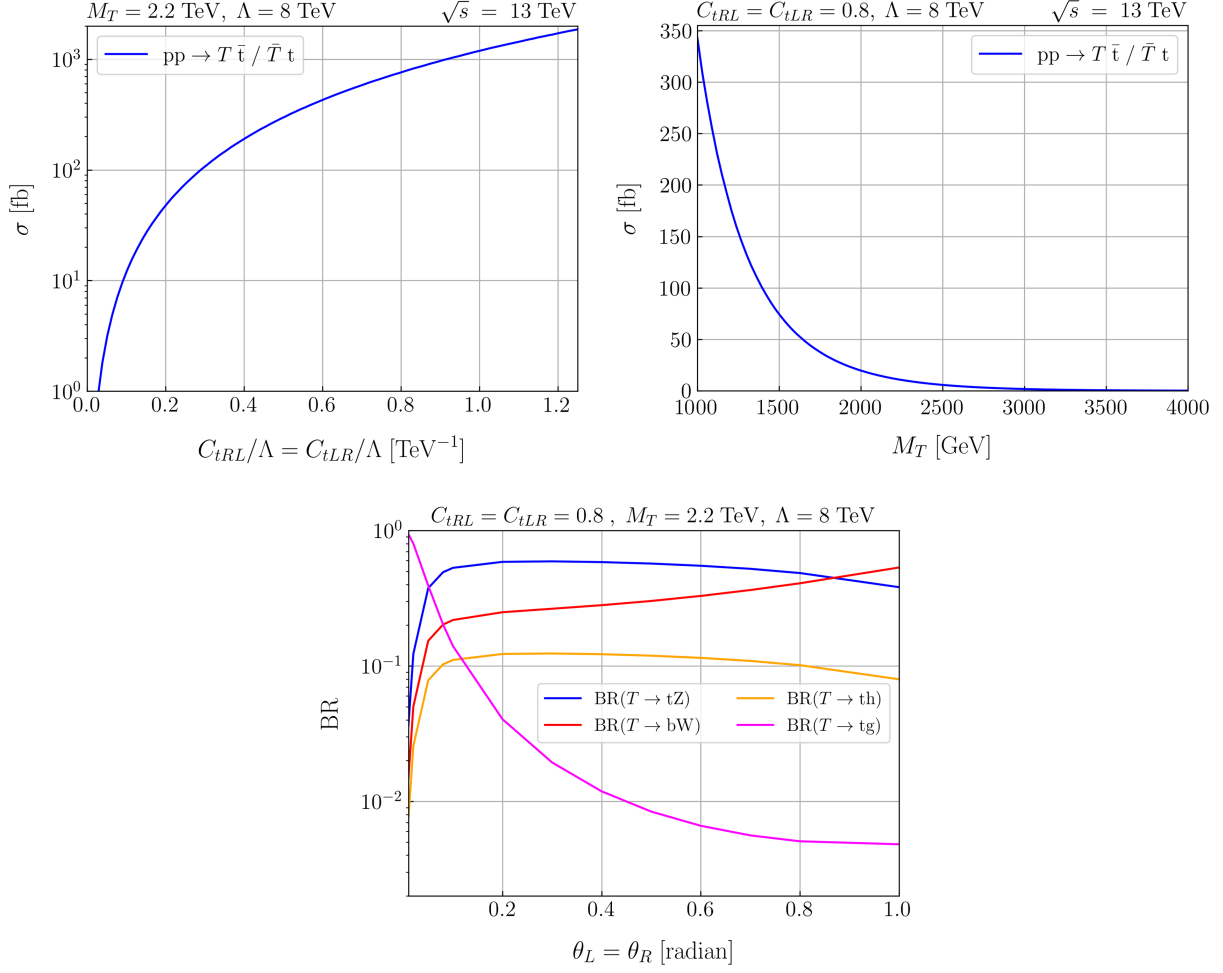


FIG. 2. Cross section vs coupling (upper left) and vs T quark mass (upper right) are presented for $pp \rightarrow T\bar{t}/\bar{T}t$ process at $\sqrt{s} = 13$ TeV with new physics scale $\Lambda = 8$ TeV, $M_T = 2.2$ TeV (upper left), and $C_{tRL} = C_{tLR} = 0.8$ (upper right). Branching ratios (BRs) vs mixing angle for all T decay modes (lower) are shown with $M_T = 2.2$ TeV, $C_{tRL} = C_{tLR} = 0.8$, and $\Lambda = 8$ TeV.

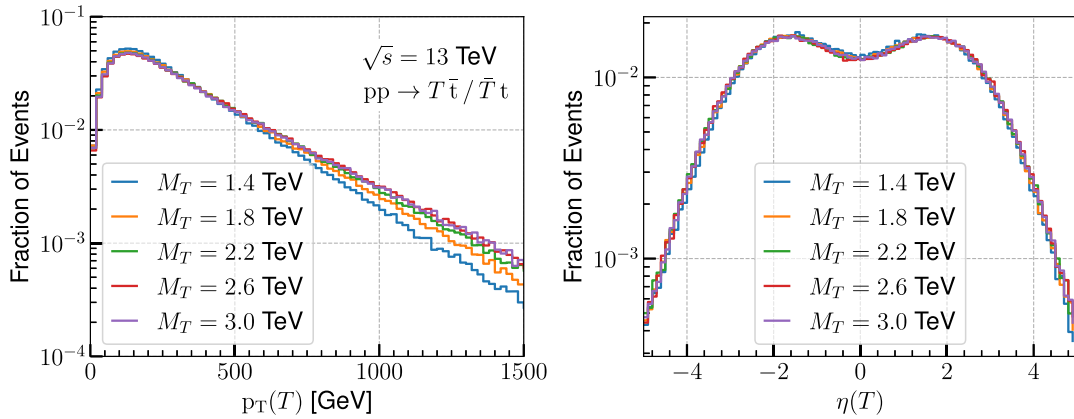


FIG. 3. Distributions of the top partner p_T (left) and η (right) at parton level for different mass hypotheses for the process $pp \rightarrow T\bar{t}/\bar{T}t$ at $\sqrt{s} = 13$ TeV.

TABLE I. Details of signal simulation tools and associated settings.

BSM model	SM + Eqs. (1)–(4) using FeynRules [62,63] and UFO [64]
Matrix element	Eqs. (5) and (6) using MadGraph5_aMC@NLO [58] at LO with $\sqrt{s} = 13$ TeV
PDF set	NNPDF30_nlo_nf_5_pdfas [65]
Parton showering and hadronization	Pythia8 [60]
Underlying event	CMS tune MonashStar [68], based on Monash 2013 tune [69]
Detector simulation model	Delphes_v3.5.0 [61]

TABLE II. Simulated sample sizes (N_{Events}) for the signal process in different T quark decay modes, along with their respective contributions to each event category as defined in Eqs. (8) and (9).

	Decay mode			
	$T \rightarrow tZ$	$T \rightarrow th$	$T \rightarrow tg$	$T \rightarrow bW$
N_{Events} (in millions)	1.0	1.0	1.0	0.6
Two fat jets category	73.00%	72.03%	76.11%	65.89%
One fat jet category	27.00%	27.97%	23.89%	34.11%

rapidly falling tail at higher transverse momenta. The higher tail of the p_T spectrum of the T quark has a distinct dependence on its mass. For the process $pp \rightarrow T\bar{t}/\bar{T}t$, the transverse momentum of the heavy top partner remains well below the cutoff scale $\Lambda = 8$ TeV throughout the analyzed region, as shown in the left panel of Fig. 3. Thus, the effective field theory description remains valid across the entire analyzed phase space region in this work. The cross section of the signal process, the T quark decay width, and its branching fractions for all of the four decay modes are listed in Appendix A for different mass hypotheses.

The generated events are then passed through Pythia8 [60] for parton showering and hadronization. In the Pythia8 simulation, the CMS tune MonashStar [68], based on the Monash 2013 tune [69], has been used to better match CMS data, especially to model the underlying event, initial-state radiation, and beam remnants. We then incorporated the effects of pileup in the signal MC samples by overlaying minimum-bias event samples with the hard-scattered events with approximately 20 average pileup interactions per event. Finally, a fast detector simulation is performed on the pileup overlaid samples using Delphes_v3.5.0 [61] that employs an overall detector geometry definition and a parametric model of average detector response.² The signal event simulation details are summarized in Table I.

Depending on the decay mode of the top partner into tZ , th , tg , or bW , the above processes could be realized into different event categories. To suppress a huge QCD multijet background at the LHC, we focus on the semileptonic

channels requiring one lepton ($\ell = e, \mu$) in the final state. Therefore, the final states where one of the W 's originating from either of the $t, \bar{t}, T$, or $\bar{T}$ decay leptonically are studied. The event sample for our search can be broadly classified into the following two categories:

$$\text{Two fat jets category: } 1\ell + 2 \text{ fat jets} + 1b - \text{jet} + p_T^{\text{miss}}, \quad (8)$$

$$\text{One fat jet category: } 1\ell + 1 \text{ fat jet} + \geq 1b - \text{jet}(s) + p_T^{\text{miss}}. \quad (9)$$

Table II lists the different decay modes of the top partner and their possible contribution to each event category, applying the same selection criteria as defined in Sec. IV on the lepton and jets. The signal fractions are larger in the two fat jets event category for all decay modes. However, in this event category, the $t\bar{t} + X$ ($X = W, Z, \text{jets, etc.}$) backgrounds also have a significantly larger contribution, leading to a poor signal-to-background ratio and thus making it unsuitable as a search region for the signal process. Therefore, we focus on the one fat jet category where the contribution of the bW decay mode is relatively higher.

C. Background processes

Several SM processes that constitute the same final state as that of the signal are considered as backgrounds. We list the relevant SM background processes for our study in Table III. The background MC samples are obtained from the CMS Open Data portal [70–76], while the cross sections are obtained from the available predictions at the highest perturbative order, with the exception of the QCD multijet process, where LO predictions are used. The accuracies of the respective cross section predictions are also given, along with the number of MC events taken for the analysis, in the same table. Among the MC samples, the dominant $t\bar{t} + \text{jets}$ events were generated at NLO accuracy with up to two additional partons using MadGraph5_aMC@NLO [58] and FxFx merging [77]. For this background process, MC samples binned in H_T are stitched with the so-called inclusive sample according to the method described in Ref. [78] for better statistical modeling of the boosted topology. Table IV summarizes the stitching weights applied to events lying in different H_T bins. The

²The choice of using Delphes fast detector simulation for signal MC samples was made owing to limitations in the available computing and storage resources at our disposal for this study.

TABLE III. List of relevant SM background processes along with their respective cross sections and MC sample sizes (N_{Events}). All MC samples are obtained from the CMS Open Data portal. The “cross section accuracy” refers solely to the order of the cross section calculation used for normalization correction, while “event generation accuracy” refers to the accuracy of the event generation at the matrix element level. References in the “process” column point to the official MC samples from the CMS Open Data portal, while those in the “cross section” column refer to the respective calculations. Here, NLO, NNLO, and NNLL denote next-to-leading order, next-to-next-to-leading order, and next-to-next-to-leading logarithmic accuracy, respectively, while EW stands for electroweak corrections.

Process	Event generation accuracy	Cross section (pb)	Cross section accuracy	N_{Events} (million)
$t\bar{t}$ + jets [70]	NLO [58,77,88]	833.900 [90]	NNLO + NNLL	117.70
$tW/\bar{t}W$ + jets [71]	NLO [79,82,88]	79.300 [91]	NNLO	2.00
$t\bar{t}Z$ [72]	LO [58,85,87,88]	0.863 [92]	NLO + NNLL	22.16
$t\bar{t}W$ [73]	NLO [58,77,88]	0.745 [93]	NNLO (QCD) + NLO (EW)	0.83
WZ + jets [74]	NLO [58,77,88]	56.850 [94]	NNLO	13.19
WW + jets [75]	NLO [82,84,88]	1.371 [95]	NNLO	8.92
QCD multijet [76]	LO [88]	1.981×10^9	LO	97.19

TABLE IV. Stitching weights for $t\bar{t}$ + jets events in different H_T bins for 300 fb^{-1} integrated luminosity, computed following the prescription in Ref. [78].

	H_T bins (GeV)				
	[0, 600)	[600, 800)	[800, 1200)	[1200, 2500)	≥ 2500
Stitching weights	2.562	2.550×10^{-2}	1.413×10^{-2}	9.180×10^{-3}	2.154×10^{-4}

event samples for WZ + jets and $t\bar{t}W$ processes were also generated at NLO accuracy with at most one extra parton using MadGraph5_aMC@NLO [58] and FxFx merging [77]. The $tW/\bar{t}W$ + jets were simulated at NLO [79] using POWHEG-BOX [80–82] with a five flavor scheme [83], while the event samples for the WW + jets process were produced at NLO accuracy [84] using POWHEG-BOX-v2 [80–82]. The $t\bar{t}Z$ process was generated using MadGraph5_aMC@NLO at LO with up to one additional jet, employing MLM merging scheme [85–87]. The QCD multijet background was generated at LO using the Pythia8 [88] generator. A combination of different QCD MC samples in nonoverlapping $\hat{p}_T$ bins above 15 GeV is considered. In all cases, NNPDF30_nlo_as_0118 [65]³ was chosen to be the default PDFset, and Pythia8 [88] was used for parton showering and hadronization; while CMS tune Monash Star [68] implemented in Pythia8 is used to model the underlying event. The effect of pileup was simulated in the background MC samples by overlaying minimum-bias events with the hard-scattered events with an average of around 20 pileup interactions per event. All background MC samples undergo full simulation of the CMS detector based on Geant4 [89] that comprises detailed geometry, alignment, and calibration information at the subdetector level.

³The PDF predictions for signal and background MC samples use exactly same number of quark flavors (5) and nominal value of the strong coupling constant at the Z pole (0.118).

III. JET ALGORITHM AND SUBSTRUCTURE METHODS

A. Dynamic radius jet clustering algorithm

The jet clustering algorithms are crucial steps in the analysis of events in a hadron collider experiment such as the LHC [29,30]. The currently used algorithms are fixed-radius sequential recombination inclusive jet algorithms, namely, the k_T (KT), anti- k_T (AK), and Cambridge-Aachen (CA) [96–100] algorithms. These traditional algorithms take as input the list of four momenta of the particles from an event and successively recombine them into several jets. In their iterative steps, these algorithms calculate pairwise k_T distance (d_{ij}) between i th and j th four momenta and “beam” distance (d_{iB}) for the i th four momentum, defined as

$$d_{ij} = \min(p_{T,i}^{2n}, p_{T,j}^{2n}) \frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \quad (10)$$

$$d_{iB} = p_{T,i}^{2n} \quad (11)$$

to decide whether to merge two four momenta or declare a particular four momentum as a final jet. Here, ΔR_{ij} is the distance between i th and j th four momenta in the pseudorapidity-azimuth (η - ϕ) plane. The two four momenta are merged to give a new four momentum if d_{ij} becomes the least among all the pairwise k_T and beam distances. Otherwise, the i th four momentum corresponding to the least beam distance is declared as a final jet. Importantly, the algorithm

has parameters n , which usually takes values -1 , 0 , and 1 corresponding to AK, CA, and KT algorithms, respectively, and R as the distance measure in the η - ϕ plane. These algorithms typically output jets of a fixed size, which is determined by the input parameter R to the algorithm. In this article, the AK algorithm is chosen as the default since AK jets have a regular conical shape with a well-defined jet area [98].

The traditional jet clustering approaches mentioned above have the limitation of producing varying-sized jets in an event for a given value of R . Several interesting attempts have been made to modify the algorithms to accommodate jets of variable sizes, such as “variable R ” jets [101], “fuzzy” jets [102], SHAPER [103], X Cone [104], PAIRed jets [105], dynamic-radius jets [35], etc. Here, we focus on the dynamic-radius jet clustering algorithm [35], which attempts to dynamically modify the jet radius during the evolution of each jet through the iterative steps of the algorithm. The approach is to start from an initial radius R_0 , the radius for each evolving “protojet” gets modified depending on its internal structure during the process of clustering. The square of the additive modifier σ_i to R_0 for each evolving protojet at each interactive stage is given by

$$\sigma_i^2 = \frac{\sum_{a<b} p_{T,a} p_{T,b} \Delta R_{ab}^2}{\sum_{a<b} p_{T,a} p_{T,b}} - \left(\frac{\sum_{a<b} p_{T,a} p_{T,b} \Delta R_{ab}}{\sum_{a<b} p_{T,a} p_{T,b}} \right)^2, \quad (12)$$

where a and b run over the constituents of the protojet. So, the final radius for the i th jet becomes $R_d = R_0 + \sigma_i$. The advantage one gets in this method is that the radius is no longer fixed; it gets adapted to the requirement of the jet that is being formed. The radius modifier can also be expressed in terms of the ECF observable $C_1^{(\beta)}$ [37] as

$$\sigma_i^2 = C_1^{(2)} - \left(C_1^{(1)} \right)^2. \quad (13)$$

For a splitting characterized by the opening angle (θ) between two hard cores, or between a single hard core and the dominant emission, the correlator behaves as $C_1^{(\beta)} \simeq \frac{E_2}{E_1} \theta^\beta$, where E_1 and E_2 denote the energies of the hardest core and the associated emission, respectively. Consequently, the radius modifier scales as $\sigma_i \propto \theta$ and is therefore directly sensitive to the decay or splitting opening angle. Furthermore, the two-point energy correlator behaves as $\sigma_i^2 \sim C_1^{(2)} \sim \frac{m^2}{p_T^2}$ for the narrow QCD jets, where emissions are nearly collinear [106]. As a result, this radius modifier adapts according to the characteristic behavior of corresponding jet types, making this clustering approach particularly effective. Thus, as expected, the radius modifier tends to become larger for top or heavy boson jets [107]; while, on the other hand, it tends to be smaller for narrow jets like the ones appearing from light quarks or gluons. Since

our study involves one fat jet along with narrow b jets, the efficiency of the DR jet clustering is expected to be better.

B. Jet substructure and parameter choices

Improving signal sensitivity over the backgrounds using jets and their substructures is the primary aim of this work. We present here a brief overview of the various jet substructure methods that are used in our analysis.

- (i) Soft drop grooming: To mitigate contamination from pileup, underlying events, etc., inside a jet, and improve the reliability of the jet substructure measurements, we have used the soft drop declustering method [51]. This method primarily aims to remove soft and wide-angle radiations inside a jet while retaining the hard core of the jet. To briefly describe, the soft drop algorithm is applied on a jet that has been reclustered using the CA algorithm (which orders by angle). The essential step of this method is to remove soft and wide-angle subjets after declustering the jet into two subjets (labeled as i and j) by checking the following condition:

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}} < z_{\text{cut}}^{\text{SD}} \left(\frac{\Delta R_{ij}}{R^{\text{SD}}} \right)^{\beta^{\text{SD}}}. \quad (14)$$

Here, $p_{T,i}$ and $p_{T,j}$ are the transverse momenta of the subjets, and ΔR_{ij} is the distance between them in the η - ϕ plane. This process is repeated until the condition fails to be satisfied. The radius R^{SD} is the characteristic jet radius usually taken to be the radius parameter used during the clustering. In the case of the DR jet clustering algorithm, we have taken this parameter to be $R_d = R_0 + \sigma$, the final dynamically adjusted radius for each jet. The two parameters of soft drop are (1) $z_{\text{cut}}^{\text{SD}}$, the energy fraction threshold, and (2) β^{SD} , the exponent regulating the angular separation between the two subjets. In the special case of $\beta^{\text{SD}} = 0$, the method reduces to the modified mass drop tagger [108,109]. The grooming then depends only on the momentum fraction z and not on the angular separation. We have used the value of $\beta^{\text{SD}} = 0$ to keep our analysis aligned with the available jet substructure measurement [107]. We also check the alternative value of $\beta^{\text{SD}} = 1$. For the energy fraction threshold parameter, a nominal value of $z_{\text{cut}}^{\text{SD}} = 0.1$ is used to match the studies performed in Refs. [51,110–112].

- (ii) N subjettness: The jet substructure method N subjettness [52], provides a measure of the consistency of a jet having N subjets inside. In our event topology, we expect boosted top, W , or Z fat jet, which gives rise to a three- or two-prong structure. Therefore, this method becomes particularly useful with the values of $N = 3$ and $N = 2$.

The N -subjettiness observable captures these subjects by fitting N subjet axes inside the fat jet. This is performed by minimizing the N -subjettiness measure [113] defined as

$$\tilde{\tau}_N^{(\beta^{NS})} = \frac{1}{d_0} \sum_{i \in \text{jet}} p_{Ti} \min \left\{ \Delta R_{i,a_1}^{\beta^{NS}}, \Delta R_{i,a_2}^{\beta^{NS}}, \dots, \Delta R_{i,a_N}^{\beta^{NS}} \right\}, \quad (15)$$

where the sum runs over all constituents in the jet, and $\Delta R_{i,a_k}$ is the distance in the η - ϕ plane between constituent i and subjet axis a_k . The prefactor $1/d_0$ is a normalization factor. The minimization is performed over the choices of axes, and the final minimum value of $\tilde{\tau}_N$, denoted as τ_N , represents the optimal measure of N subjettiness. The exponent β^{NS} and the method used to determine the seed N axes (e.g., exclusive k_T axes from an iterative algorithm) are parameters of choice. We have used the N -subjettiness method implemented in Fastjet Contrib [114]. In our study, we use $\beta^{NS} = 1$, the seed axes obtained by the WTA_KT_Axes method [115–117], and the ratios $\tau_{32} = \tau_3/\tau_2$ and $\tau_{21} = \tau_2/\tau_1$ to, respectively, identify three- and two-prong structures inside a fat jet.

- (iii) Energy correlation functions: These are a class of infrared and collinear safe observables proposed to study jet properties [37]. The first three lowest point functions take the form

$$\text{ECF}_1(\beta^{\text{ECF}}) = \sum_{i \in \text{jet}} p_{Ti}, \quad (16)$$

$$\text{ECF}_2(\beta^{\text{ECF}}) = \sum_{i < j \in \text{jet}} p_{Ti} p_{Tj} \Delta R_{ij}^{\beta^{\text{ECF}}}, \quad (17)$$

$$\begin{aligned} \text{ECF}_3(\beta^{\text{ECF}}) = & \sum_{i < j < k \in \text{jet}} p_{Ti} p_{Tj} p_{Tk} \\ & \times (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^{\beta^{\text{ECF}}}, \end{aligned} \quad (18)$$

where the sum is over the jet constituents and the exponent β^{ECF} is a parameter that gives the weight to the angular separation between the jet constituents. These functions help distinguish between jets originating from different partonic configurations. In the case of top quark study, a better discrimination from QCD jets is obtained for $\beta^{\text{ECF}} = 1.0$ [37]. We therefore choose to use $\beta^{\text{ECF}} = 1.0$ as a nominal value and $\beta^{\text{ECF}} = 0.5$ for an alternate value. Instead of using the ECFs as defined in Eqs. (16)–(18) directly, their ratios are better suited as they remove the p_T dependence. The following ratios [107] are studied:

TABLE V. Parameter choices for jet substructure methods applied to the DR fat jets. “Nominal values” correspond to those used in the final analysis, while “alternate values” are considered for the control study shown in Fig. 4.

Method	Parameter	Nominal values	Alternate values
Soft drop [51]	β^{SD}	0.0 [107]	1.0
	$z_{\text{cut}}^{\text{SD}}$	0.1 [107]	0.1
N subjettiness [52]	β^{NS}	1.0 [107]	1.0
	Seed axes [104]	WTA_KT_AXES [110]	WTA_KT_AXES
ECF [37]	β^{ECF}	1.0 [107]	0.5

$$e_2 = \frac{\text{ECF}_2}{(\text{ECF}_1)^2}, \quad e_3 = \frac{\text{ECF}_3}{(\text{ECF}_1)^3}, \quad (19)$$

$$C_2 = \frac{e_3}{(e_2)^2}, \quad \text{and} \quad D_2 = \frac{e_3}{(e_2)^3}. \quad (20)$$

The DR jet algorithm, along with the jet substructure methods, has been implemented within Delphes and CMS software version 7_6_7 as appropriate plug-ins. The choices of parameters for the jet substructure methods described above for the DR jets, including both nominal and alternate values used for variations, are summarized in Table V. The impact of these choices is studied by comparing the shapes of different observables sensitive to the respective substructure methods, as shown in Fig. 4, both at truth and reconstructed levels. As seen in the figure, the distributions for nominal and alternate parameter choices look quite similar, suggesting that the overall analysis remains fairly robust against these variations. In addition, substructure observables of the DR jets are also compared with the fixed-radius counterpart at the reconstructed level and are shown in Fig. 5. Only the nominal values of the substructure parameters are considered for this comparison.

The upper left panel of Fig. 5 shows the distribution of soft drop mass (m_{SD}) for both fixed- and dynamic-radius algorithms. The upper right panel in Fig. 5 shows the variation of the resolution of the average m_{SD} of the leading fat jet with the number of vertices. From the figure, we see that the DR clustering consistently performs better than the fixed-radius method, especially at the large vertex multiplicities. The radius modifier, defined in Eq. (12), is the p_T -weighted standard deviation of the two-particle angular distance in the η - ϕ plane among the jet constituents. Thus, the impact of the particles coming from pileup vertices is partially canceled because those are uncorrelated to the particles originating from the primary vertex constituting the jet core. Figure 5 also demonstrates that the distributions of τ_{21} (lower left) and τ_{32} (lower right) for the DR

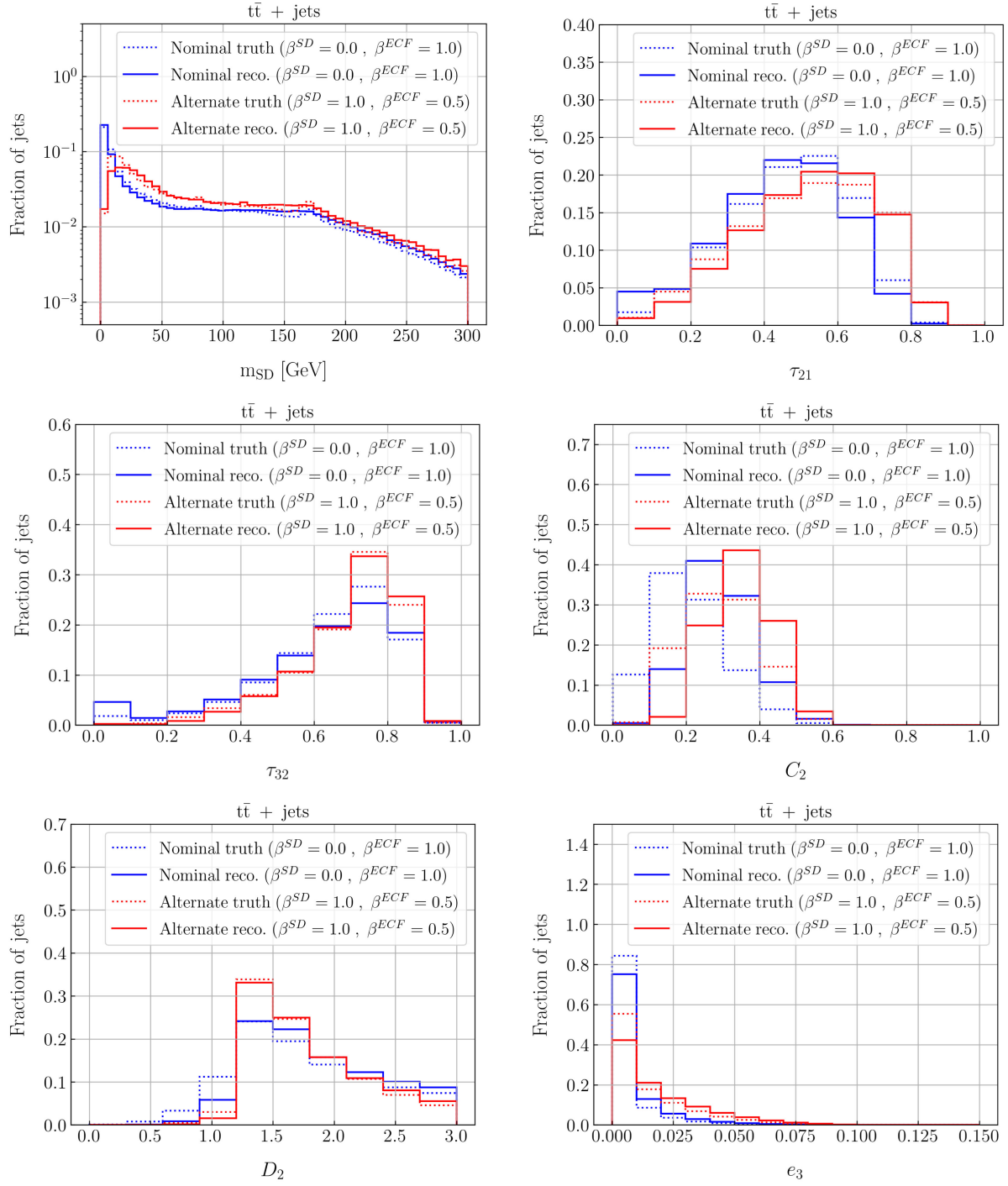


FIG. 4. Comparisons of soft drop mass (upper left), τ_{21} (upper right), τ_{32} (middle left), and the ECF ratios C_2 (middle right), D_2 (lower left), and e_3 (lower right) distributions of the leading anti- k_T fat jet corresponding to nominal and alternate values of substructure parameters as indicated in Table V for dynamic-radius jet clustering algorithm at truth and reconstructed levels in $t\bar{t}$ + jets events.

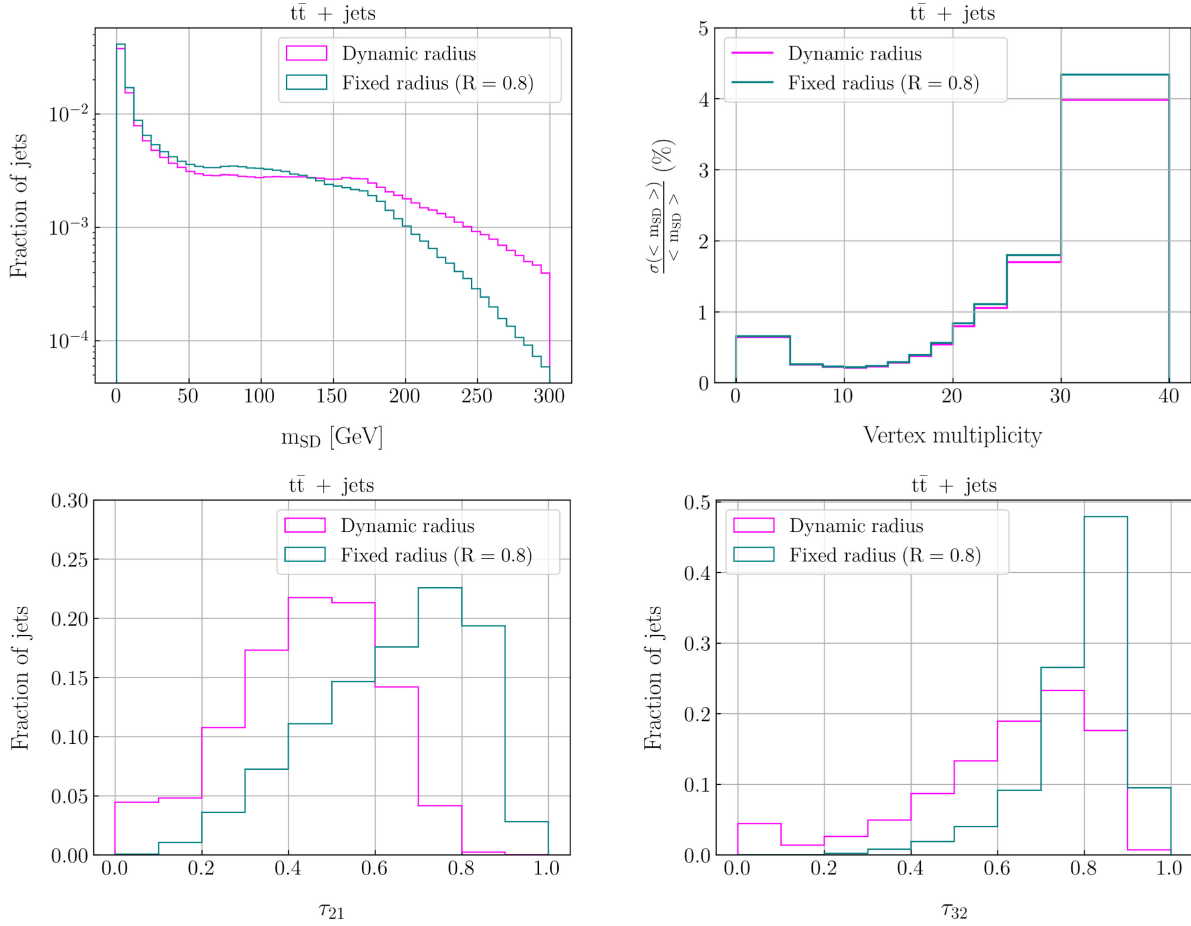


FIG. 5. Comparisons of soft drop mass (upper left) and its resolution in bins of vertex multiplicity (upper right), τ_{21} (lower left), and τ_{32} (lower right) distributions of the reconstructed leading anti- k_T fat jet between fixed-radius ($R = 0.8$) and dynamic-radius jet clustering algorithms in $t\bar{t} + \text{jets}$ events are presented. The nominal choice of substructure parameters as indicated in Table V has been used for the dynamic-radius algorithm to obtain these plots.

clustering are more shifted toward 0 compared to the fixed-radius clustering, indicating better reconstruction of the hadronically decaying boosted W bosons and top quarks in the former case with a fat jet [52].

IV. EVENT RECONSTRUCTION AND SELECTION

Various key features of the event reconstruction and selection are presented below.

- (1) Selection criteria: Events having a minimum of one reconstructed pp interaction vertex are retained for this study. The vertices must be reconstructed from at least four tracks that have a longitudinal distance $|d_z| < 24$ cm and a radial distance $|d_{xy}| < 2$ cm from the nominal interaction point. If multiple vertices are found in an event, the one with the highest value of summed p_T^2 of clustered particles is taken as the primary pp interaction vertex. In contrast, the others are identified as the pileup

vertices. Individual particles are reconstructed via the particle-flow algorithm [118]. Muons (electrons) are required to have $p_T \geq 30(35)$ GeV, $|\eta| \leq 2.4(2.1)$ and to pass the “tight” identification criteria [119,120]. The selected muon and electron candidates must pass the criterion on relative isolation (I_{rel}) to be within 6% of their respective transverse momentum. Additionally, the selected electron candidate is required to satisfy $|d_{xy}| \leq 0.05(0.1)$ and $|d_z| \leq 0.1(0.2)$ cm relative to the primary vertex in the barrel (end cap) region of the electromagnetic calorimeter (ECAL). Events containing additional muons (electrons) with $p_T \geq 10(15)$ GeV, $|\eta| \leq 2.4(2.5)$, and passing the “loose” (“veto”) identification criteria are rejected. In such cases, the criteria on the additional lepton isolation are relaxed to $I_{\text{rel}} \leq 20\%$ for muons and $I_{\text{rel}} \leq 18\%(16\%)$ for electrons in the barrel (end cap) ECAL region. Standard jets are reconstructed using the AK clustering algorithm [98] with radii of 0.4 and 0.8 for the

TABLE VI. Event yields of signal and background processes for the two lepton flavors after applying all selection criteria for an integrated luminosity of 300 fb^{-1} .

Final state	Background processes						Total bkg.	Signal
	$t\bar{t} + \text{jets}$	$tW/\bar{t}W + \text{jets}$	$t\bar{t}Z$	$t\bar{t}W$	$WV + \text{jets}$ ($V = W \text{ or } Z$)	QCD multijet		
$e + \text{jets}$	18313	29544	620	825	4077	35634	89013	25
$\mu + \text{jets}$	69083	99413	2112	3035	13243	63309	250196	15

narrow (AK4) and fat (AK8) jets, respectively, as implemented in the `FastJet` package [114] with the particle-flow candidates as inputs. The effect of additional tracks and energy deposits in calorimeters from pileup on the AK4 and AK8 jet momenta is mitigated by discarding tracks identified to be originating from pileup vertices, as well as by applying an offset correction to account for residual neutral pileup contributions [121,122]. “Tight lepton veto” identification criteria [123] are applied to suppress jets arising from spurious sources, such as electronic noise in the calorimeters and jets containing tightly identified muons or electrons. The combined secondary vertex tagging algorithm version 2 (CSVv2) [124] is used to identify AK4 jets originating from bottom quarks or antiquarks. Candidate events are required to have exactly one AK8 jet with $p_T \geq 200 \text{ GeV}$ and within $|\eta| \leq 4.0$, as mentioned in Sec. II B. In addition to the selected AK8 jet, events must have at least one nonoverlapping AK4 jet with $p_T \geq 30 \text{ GeV}$, $|\eta| \leq 2.5$ and passing the “medium” working point defined on the CSVv2 score. Finally, events should pass the $p_T^{\text{miss}} \geq 30 \text{ GeV}$ criterion.

We note that a full detector simulation with `Geant4` with CMS detector specification has been used in the background MC samples. However, for the signal MC samples, a fast detector simulation using `Delphes` has been used. According to Ref. [61], the overall agreement in the average performance of `Delphes` with the full simulation of the CMS detector based on `Geant4` is reasonably good for different high-level physics objects, especially in the phase space region of interest. Furthermore, the measured efficiencies of lepton (both selected and vetoed leptons) [119,120] and jet [123,124] reconstruction are incorporated into `Delphes` for signal events in bins of their respective p_T and η . In case of backgrounds, to mitigate observed differences with collision data, respective efficiency corrections are applied [119,120,123,124]. Table VI lists signal and background yields after all event selection steps for the two lepton flavors, while Figs. 6–8 show the distributions of observables related to lepton, jet, and event kinematics as indicated in Table VIII.

- (2) Reconstruction of dynamic-radius jets: In addition to the AK4 and AK8 jets, we reconstruct the DR anti- k_T jets (see Sec. III A) for the selected events⁴ using the particle-flow candidate information as input. The DR algorithm is used to form two jet collections per event, corresponding to R_0 values of 0.3 and 0.7 for narrow and fat jets, respectively. The narrow jets that do not overlap with any fat jet in the event are considered for further analysis. The CSVv2 score of the nearest AK4 jet is assigned as the b-tagging discriminant value for each narrow jet.
- (3) Signal event reconstruction with dynamic-radius jets: It is important to understand the impact of the DR clustering method on the signal event reconstruction, especially in the boosted regime. A simple study is performed for this purpose by applying the following selections on the jet substructure observables shown in Fig. 5, to identify jets due to hadronically decaying boosted W or Z bosons and top quarks in the simulated signal events.
 - (a) Boosted W- or Z-like jet selection criteria: $m_{SD} \in [60.0, 100.0] \text{ GeV}$, $\tau_{21} < 0.5$, and fat jet $p_T \geq 900 \text{ GeV}$.
 - (b) Boosted toplike jet selection criteria: $m_{SD} \in [150.0, 200.0] \text{ GeV}$, $\tau_{32} < 0.5$, and fat jet $p_T \geq 900 \text{ GeV}$.

The results are presented in Table VII in terms of the selection efficiencies of the boosted W or Z bosons and top quarks for the two clustering methods at high fat jet p_T . Higher selection efficiencies are observed for both the boosted W- or Z-like and toplike jets in the case of the DR clustering method, indicating better reconstruction of the daughters of the T quark in the boosted regime, thus improving the overall signal event reconstruction in this region. The adaptive approach adopted by the DR clustering (see Sec. III A) helps in capturing all daughters of the boosted W or Z boson or the top quark originating from the T quark decay (see Table II) within the fat jet by optimizing the jet size.

⁴The jet selection criteria are applied to AK4 and AK8 jets, and not to dynamic-radius jets. This choice is made primarily due to the absence of optimized pileup contamination removal, jet identification working points, jet energy resolution and correction, etc., for the dynamic-radius jets.

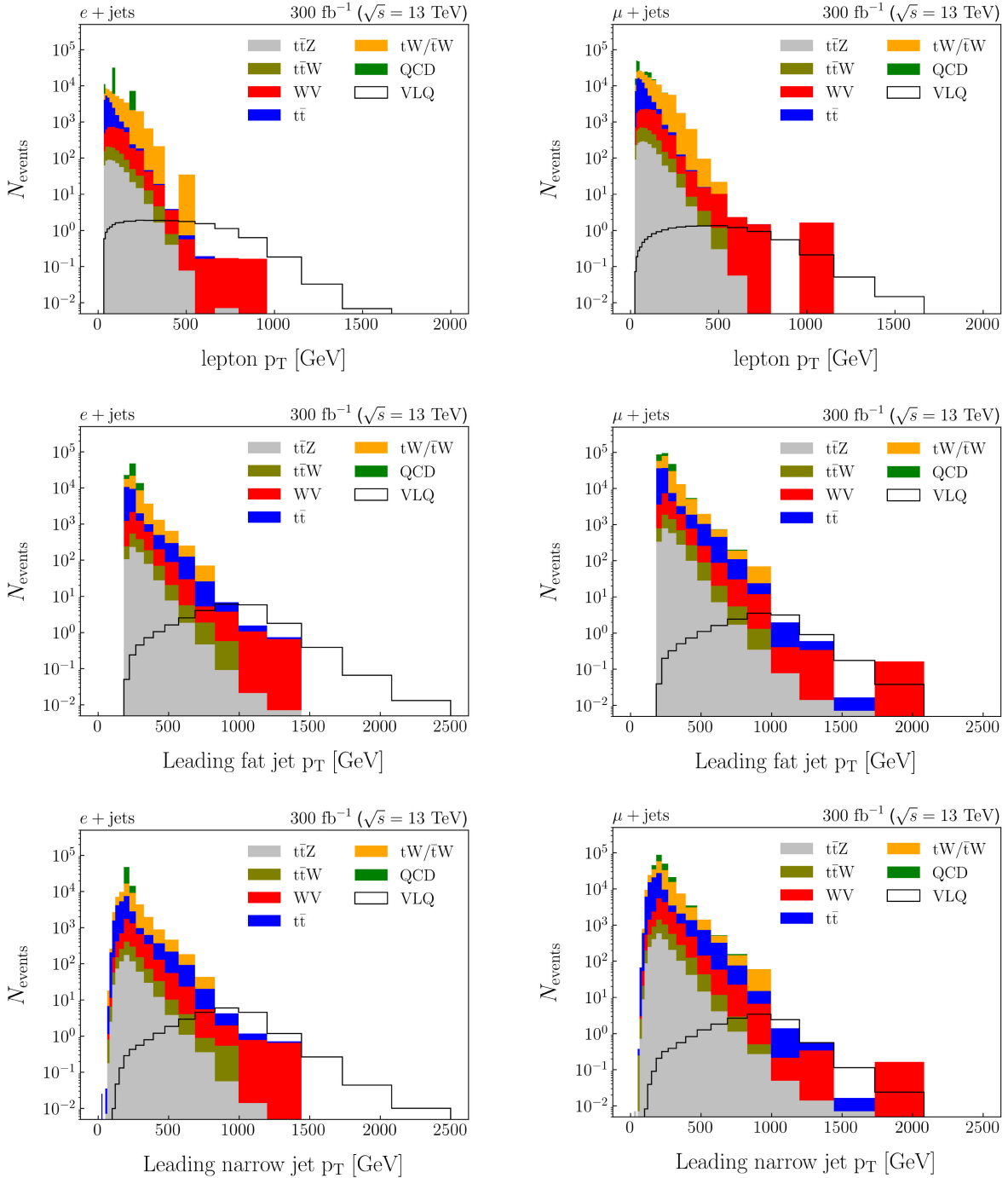


FIG. 6. Distributions of lepton p_T (upper), fat jet p_T (middle), leading narrow jet p_T (lower), in the $e + \text{jets}$ (left) and $\mu + \text{jets}$ (right) final states. In the legends, VLQ denotes the signal process with a vectorlike top partner quark, while the others correspond to the SM backgrounds as given in Table VI.

V. RESULTS

A. Multivariate analysis

Several observables are combined into boosted decision tree (BDT) discriminants to effectively separate signal from the collective SM backgrounds listed in Table III. The input observables to the BDT discriminants are listed in

Table VIII. The most important aspects of the BDT analysis are discussed in detail below.

- (i) Choice of inputs: A variety of jet substructure and event observables have been studied for this purpose, some of which are shown in Figs. 4–8. However, a subset of these is chosen as inputs to the BDT discriminants since there are large correlations

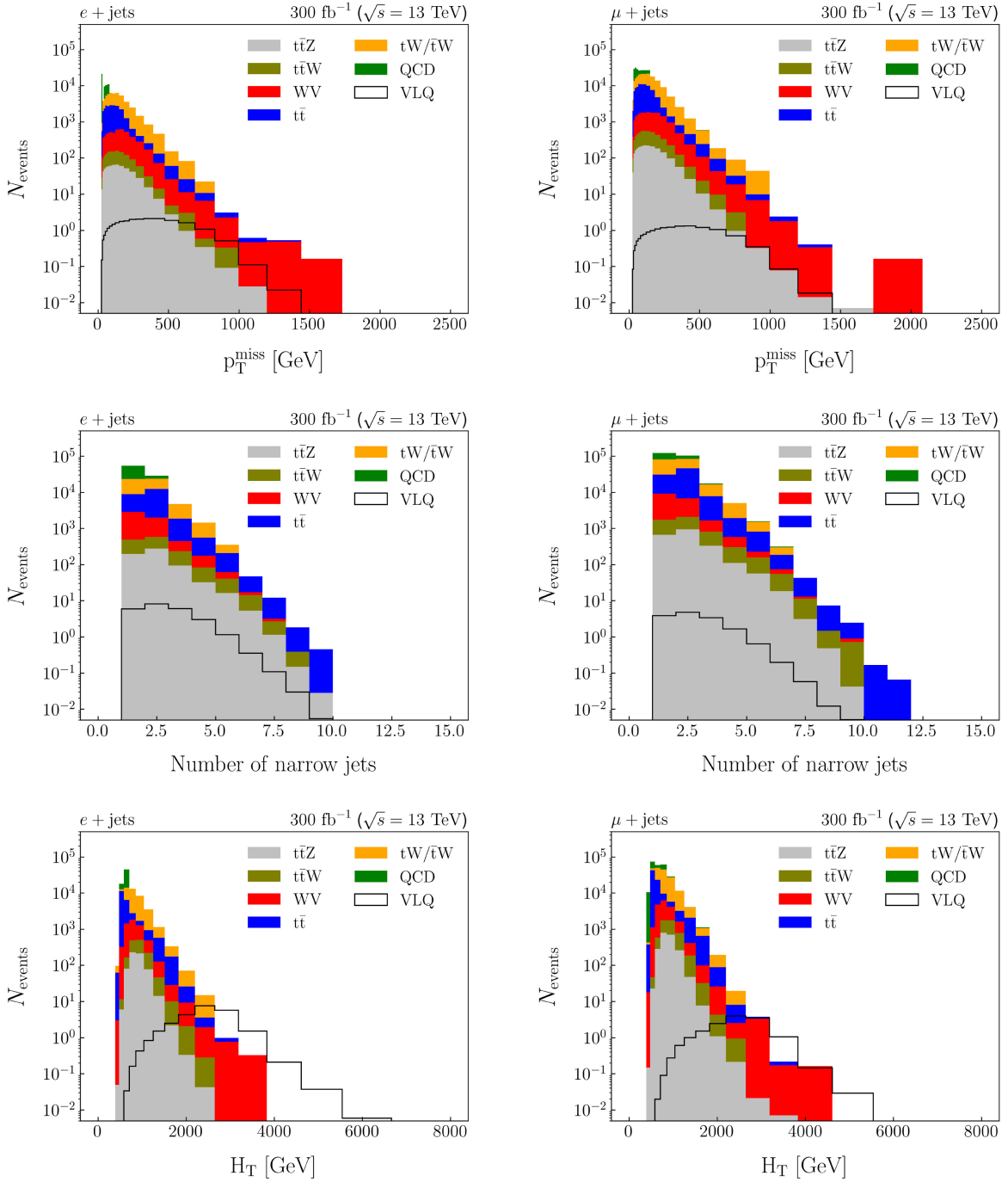


FIG. 7. Distributions of p_T^{miss} (upper), number of narrow jets (middle), and H_T (lower) in the $e + \text{jets}$ (left) and $\mu + \text{jets}$ (right) final states. In the legends, VLQ denotes the signal process with a vectorlike top partner quark, while the others correspond to the SM backgrounds as given in Table VI.

between some of the pairs. Observables that were highly correlated with each other were removed from the list of inputs to keep only the less correlated ones. For example, none of the ECFs or their ratios have been used as the inputs to the BDT discriminants because they would be highly correlated with N subjettness and their ratios. Different observables have different capabilities of discriminating the

signal process from the SM backgrounds, which can be broadly classified into two types, namely, reducible and irreducible. The backgrounds having top quarks in the final state compose the irreducible type, while the others constitute the reducible type. Boosted decision trees are fully tree-based, human-readable structures with a finite maximum depth that selects per tree only a few from the input observables

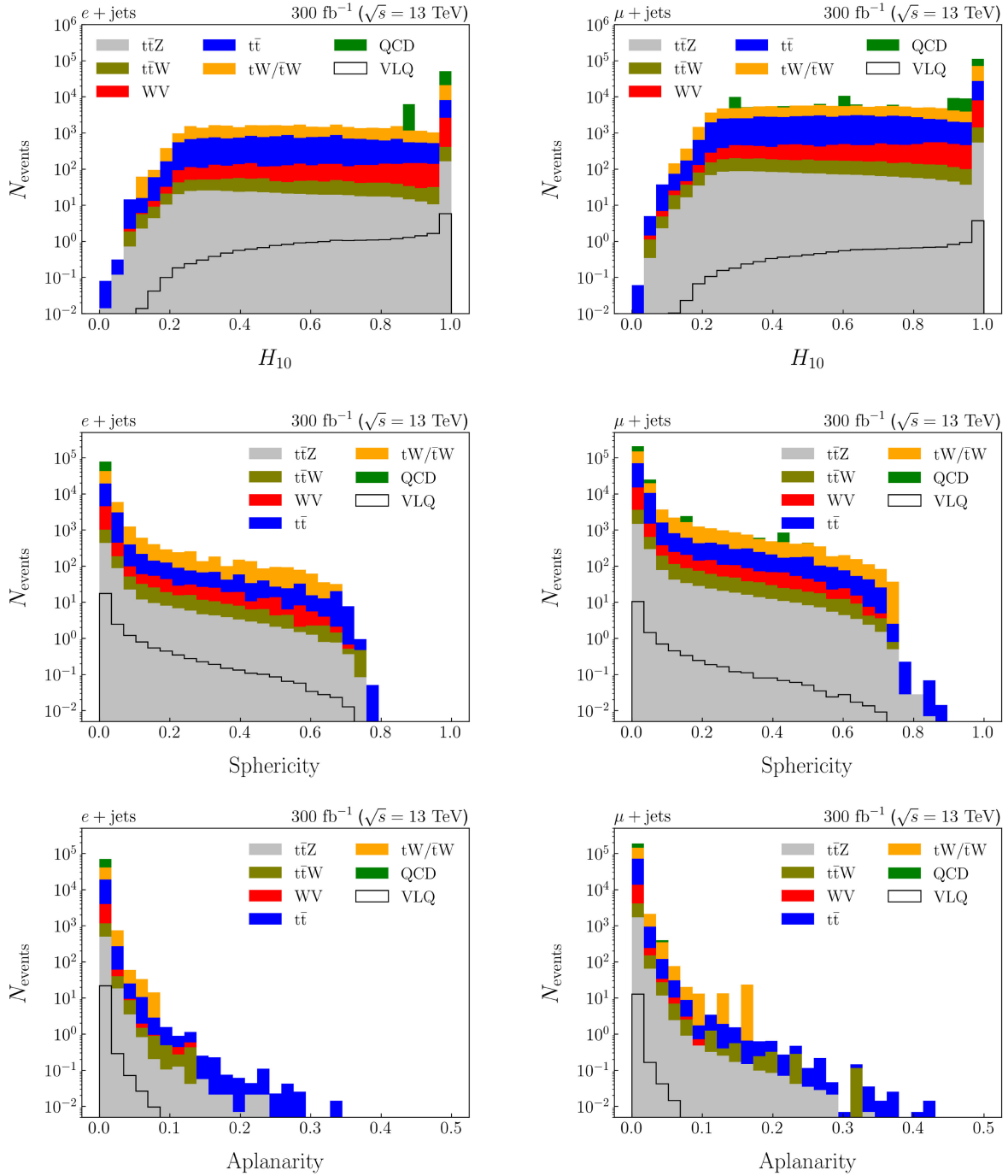


FIG. 8. Distributions of event shape observables, namely, tenth-order Fox-Wolfram moment (upper), sphericity (middle), and aplanarity (lower), in the $e + \text{jets}$ (left) and $\mu + \text{jets}$ (right) final states. In the legends, VLQ denotes the signal process with a vectorlike top partner quark, while the others correspond to the SM backgrounds as given in Table VI.

list, which provides the best discrimination between signal and backgrounds. Thus, using a relatively large number of observables as inputs to the BDTs does not introduce any bias; instead, it mostly increases the computational cost due to the residual correlations among the observables, even in the worst case where the performance is not significantly improved by including an observable. This strategy has been

widely adopted in experimental searches such as Ref. [49] as well as in phenomenological studies such as Ref. [40]. The input observables are ranked according to their so-called “importance” to the BDT discriminants as shown in Appendix B. The importance of an observable is derived by counting how often that observable has been used to split decision tree nodes and by weighting each split

TABLE VII. Selection efficiencies of boosted W- or Z-like and toplike jets for fat jet $p_T \geq 900$ GeV corresponding to fixed- and dynamic-radius clustering methods in simulated signal events.

Process	Clustering method	Selection efficiency	
		Boosted W or Z boson (%)	Boosted top quark (%)
Signal	Fixed radius	38.3	46.5
	Dynamic radius	41.1	47.5

occurrence by the separation gain squared it has achieved and by the number of events in the node [125].

- (ii) Training setup: The simulated events are separated into two independent subsamples, one for training and the other for validation, in 70%:30% ratio. An event is picked randomly to belong to any one of these subsamples. The adaptive boosting algorithm [126] implemented within the TMVA package [127] has been used to train the discriminants, keeping the learning rate, minimum node size, and maximum depth set to 0.5, 2.5%, and 3, respectively. Two independent trainings are performed, one for fixed-radius jets and another for DR jets, with exactly the same settings. In each training, samples are selected after applying a p_T threshold of 200 GeV on the leading fat jet as indicated in Sec. IV. Each process is weighted according to its purity, defined as $\frac{N_i}{\sum_i N_i}$ with i running over the signal and background processes, during training. Checks for overtraining

are performed by comparing the respective BDT scores of signal and background between training and validation subsamples. The corresponding Kolmogorov-Smirnov probabilities [128] range between 32% and 97%, indicating no overtraining. In the left column of Fig. 9, we show the BDT score distributions of signal and background corresponding to fixed-radius (upper left) and DR (lower left) algorithms for the nominal event selection.

- (iii) Performance evaluation: We evaluate the BDT scores also for events passing different thresholds on the leading fat jet p_T , and the performances of the BDT discriminants are then studied via the receiver-operator characteristic (ROC) curves for each of these thresholds separately, along with the nominal case. This exercise is primarily aimed at understanding the sensitivity in the high fat jet p_T regime. The ROC curves corresponding to different fat jet p_T thresholds are shown in the right column of Fig. 9 for fixed-radius (upper right) and DR (lower right) clustering methods. The BDT performance falls off less rapidly for tighter thresholds on fat jet p_T in the case of DR jet clustering as indicated by the ROC curves in Fig. 9.

B. Signal sensitivity and interpretation

The BDT discriminants are then utilized for a comparison of signal sensitivities followed by interpretations in terms of expected limits on the plane spanned by T quark mass and its chromomagnetic coupling, corresponding to the two clustering approaches.

TABLE VIII. List and description of input observables to the BDT discriminants.

Type	Observable	Description
Lepton kinematics	$p_{T,\ell}$	Lepton p_T
	$ \eta_\ell $	Lepton $ \eta $
Jet kinematics	$p_{T,j}$	p_T of the leading narrow jet that is not part of any fat jet
	$ \eta_j $	$ \eta $ of the leading narrow jet that is not part of any fat jet
	m_j	Mass of the leading narrow jet that is not part of any fat jet
	$p_{T,fj}$	p_T of the leading fat jet
	$ \eta_{fj} $	Leading fat jet $ \eta $
Jet substructure	m_{SD}	Soft drop mass of the leading fat jet
	τ_{21}	2-to-1 subjettness ratio of the leading fat jet
	τ_{32}	3-to-2 subjettness ratio of the leading fat jet
Event kinematics	p_T^{miss}	Missing transverse momentum
	m_T	Transverse mass of the lepton- p_T^{miss} system
	H_T	Scalar p_T sum of nonoverlapping narrow and fat jets
	N_{jets}	Number of narrow jets that are not part of the fat jet
	$N_{b\text{-jets}}$	Number of narrow b jets that are not part of any fat jet
	H_{10}	tenth-order Fox-Wolfram moment [56,57]
	S	Sphericity obtained from the eigenvalues of the momentum tensor [55]
	A	Aplanarity obtained from the eigenvalues of the momentum tensor [55]

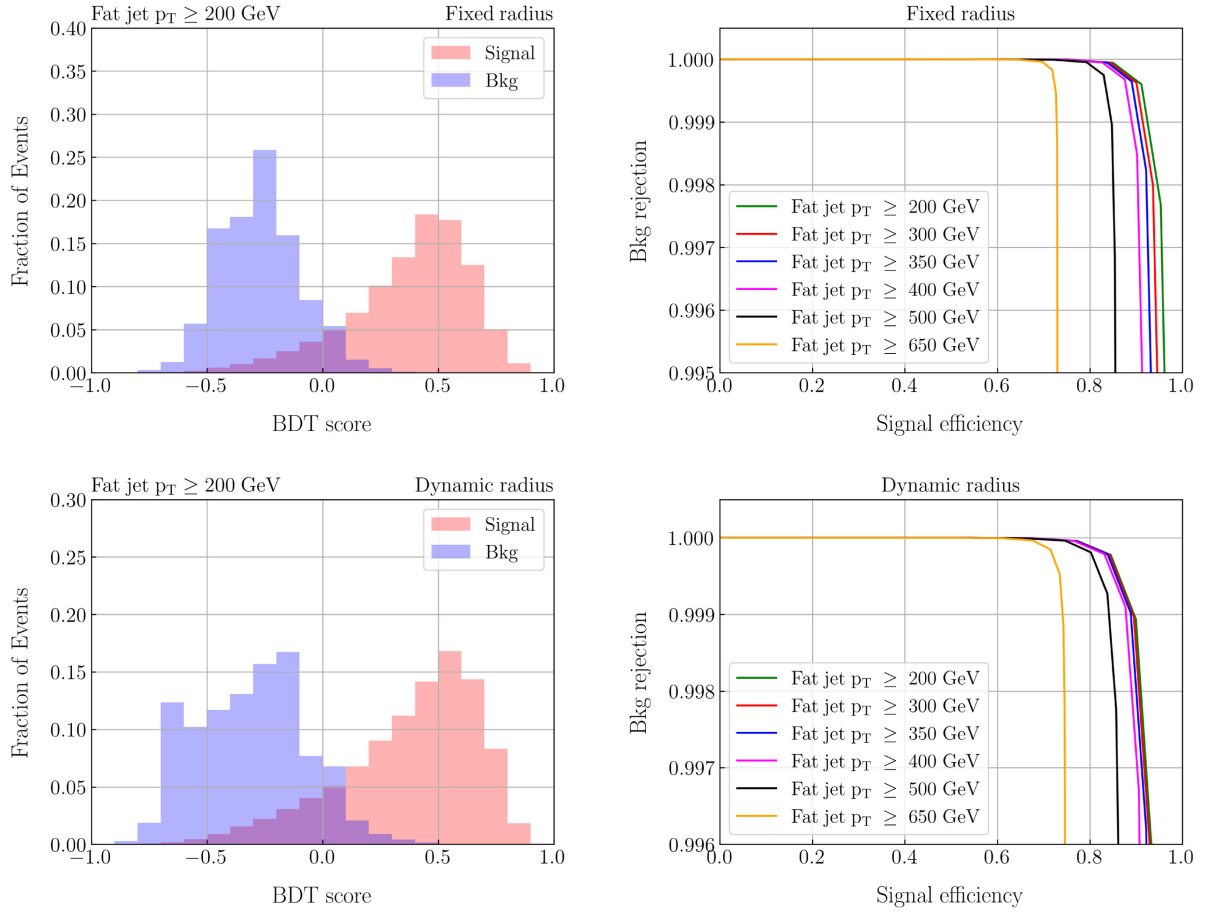


FIG. 9. BDT scores (left) along with the ROC curves (right) for fixed-radius (upper) and dynamic-radius (lower) jets. The BDT score distributions are obtained for the nominal fat jet p_T threshold of 200 GeV (see Sec. IV), which corresponds to the green curves on the right column on each row. The other ROC curves are obtained by applying an additional criterion on the fat jet p_T as indicated in the legend.

- (i) Comparative study of signal sensitivities: Selection thresholds on the BDT scores are optimized based on the so-called “Punzi figure of merit” ($\text{FOM}_{\text{Punzi}}$) defined in Ref. [129]. The signal sensitivity is characterized via two different metrics, namely, the signal significance defined as $\frac{S}{\sqrt{S+B}}$ and the $\text{FOM}_{\text{Punzi}}$, to check the consistency and robustness of the conclusion. The variation of these metrics is studied with the fat jet p_T threshold for the fixed-radius and DR jet clustering methods after applying the optimized selection on the BDT scores ≥ 0.3 . From Fig. 10, it is evident that the DR jet algorithm outperforms the fixed-radius one in the highly boosted (fat jet $p_T \geq 750$ GeV) region of the phase space. The DR algorithm shows higher sensitivity toward large M_T values (see Fig. 3) compared to the fixed-radius algorithm, allowing for more phase space to hunt for the signal process.
- (ii) Interpretation via expected limits: The result of the signal sensitivity study is interpreted by calculating

the expected 2σ upper limits on the chromomagnetic couplings, assuming $C_{iRL} = C_{iLR}$, for the top-partner masses in the range 1.4–3 TeV. For each mass point, the expected upper limit has been obtained by varying the coupling strength such that the signal significance reaches 2σ . Figure 11 shows the resulting upper limits for the three different fat jet p_T thresholds. In these plots, the shaded regions correspond to values of the coupling excluded by our analysis at 95% confidence level under the assumption of no signal observation. The upper limits are estimated for both fixed-radius and DR jet clustering methods. In all cases, the expected limit for the DR clustering is smaller than that for the fixed-radius one, indicating that the analysis with DR jets has higher sensitivity to smaller coupling values. The improvement is more prominent for the fat jet p_T threshold of 900 GeV, suggesting that the DR jet clustering is particularly effective for reconstructing signal events in the high fat jet p_T regime, as indicated earlier in Sec. IV.

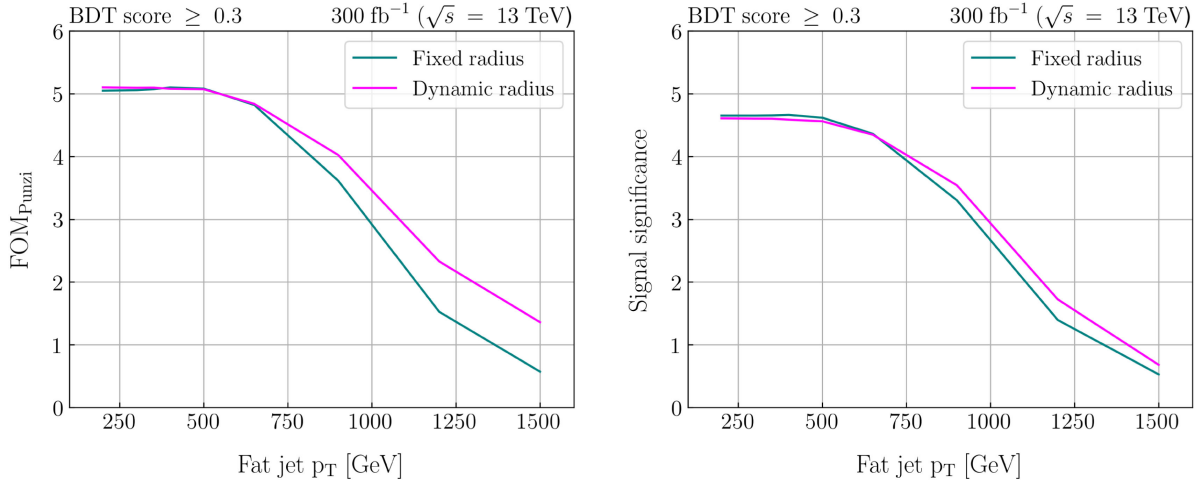


FIG. 10. Comparison of $\text{FOM}_{\text{Punzi}}$ (left) and signal significance (right) between fixed- and dynamic-radius clustering methods as a function of the fat jet p_T threshold for the benchmark scenario defined in Eq. (7).

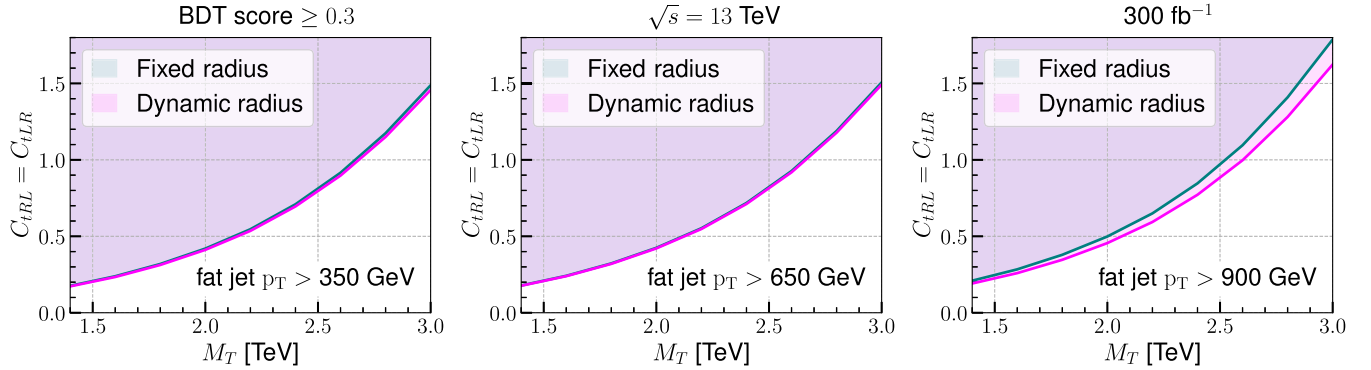


FIG. 11. Expected 2σ upper limit on the chromomagnetic coupling $C_{tRL} = C_{tLR}$ as a function of the top-partner mass. The three panels correspond to three different fat jet p_T thresholds. For these plots, we use $\Lambda = 8 \text{ TeV}$, and all other parameters are kept the same as in Eq. (7). Shaded regions indicate the parameter space excluded by our analysis, assuming no signal is observed.

C. Discussions on the dominant systematic effects

The dominant systematic effect on the signal sensitivity stems from the inaccuracy of the jet energy estimation. The jet energy is primarily determined by calorimeter measurements for high- p_T jets. The jet energy scale plays a significant role at high p_T due to partial leakage of high-energy hadronic showers beyond the longitudinal depth provided by the calorimeter that leads to an inaccurate estimation of the overall jet energy. It is likely to have a similar impact for both jet clustering methods discussed here since it is entirely dependent on the incident energy and the leakage fraction of the hadronic shower produced by the jet constituents, irrespective of the clustering method. However, the uncertainty due to jet energy resolution is expected to increase for DR clustering since more particles are likely to be accumulated within the jet, owing to its larger capture radius, thereby leading to higher fluctuations in the measured jet energy.

VI. SUMMARY AND OUTLOOK

In this work, we performed a thorough investigation of collider prospects of a vectorlike top partner produced in association with a Standard Model top quark via a chromomagnetic coupling. The top partner is assumed to decay into the following channels: tZ , $t\theta$, $t\eta$, or tW . To effectively suppress the huge QCD background, we focused on semileptonic final states, i.e., events with exactly one isolated lepton arising either from the top quark or from the heavy boson decay. Depending on the decay modes of the top partner, the final state could contain one or two fat jets along with a single lepton and b-tagged jet(s). We performed the analysis in the event category that contains a single fat jet ($1\ell + 1 \text{ fat jet} + \geq 1 \text{ b-tagged jet}$) due to its comparatively better signal-to-background ratio.

We carried out a detailed collider analysis at the $\sqrt{s} = 13 \text{ TeV}$ with an integrated luminosity of 300 fb^{-1} , incorporating realistic pileup conditions. For signal events, the

MadGraph5_aMC@NLO+Pythia8+Delphes simulation chain has been employed; while for background estimation, appropriate simulated samples from CMS Open Data are utilized. To enhance signal discrimination, we utilized state-of-the-art jet substructure tools such as soft drop grooming and N subjettiness, along with several event shape observables like Fox-Wolfram moments, sphericity, and aplanarity. Multivariate discriminants based on boosted decision trees are developed for better signal-to-background separation. We then optimized a selection threshold on these discriminants based on the Punzi figure of merit to improve signal sensitivity. The result is interpreted as the upper limit on the chromomagnetic coupling between the top quark and its heavy vectorlike partner. Assuming the cutoff scale of $\Lambda = 8$ TeV, the upper limit on the coupling is found to lie within 0.2–1.5 for the top-partner mass range of 1.4–3 TeV.

An important aspect of our work is a comparative performance analysis between conventional fixed-radius vs modern dynamic-radius clustering of anti- k_T jets under realistic pileup conditions. We observe that the dynamic-radius approach enhances the signal sensitivity, particularly in regions with high jet transverse momentum, allowing for tighter bounds on the chromomagnetic coupling. This suggests that the dynamic-radius jet clustering is more effective in probing high-energy regimes compared to the conventional fixed-radius method.

The utility of DR clustering is expected to become even more pronounced at the future high-energy colliders such as the HL-LHC and FCC-hh. At the HL-LHC, the anticipated 10 times higher integrated luminosity will significantly enhance the signal event yield in the high- $p_T(T)$ tail. Therefore, the statistical sensitivity in this phase space region is expected to improve at HL-LHC. The FCC-hh collider, which is expected to operate at the proposed center-of-mass energy of ~ 100 TeV, will further extend the reach of the high- $p_T(T)$ tail along with substantial enhancement in the partonic cross section for $pp \rightarrow Tt$ production. As a result, the heavy top partner and its decay products are expected to acquire higher transverse momenta, where the DR clustering approach is much more effective compared

to the fixed-radius approaches. These features make DR jets especially well suited for future collider environments with jets having extreme boosts. This method, when combined with advanced machine-learning-based tagging algorithms for identifying boosted heavy resonances, becomes increasingly relevant for higher sensitivity.

There remain various scopes for further improvements. Optimization of pileup contamination removal working points, as well as determining jet energy correction and resolution accurately for the dynamic-radius clustering approach, can be extremely useful. This might put even tighter bounds on the couplings than the existing ones. Future directions also include exploring the integration of DR jets with particle-based or graph neural-network-based tagging algorithms for boosted heavy resonance identification, which can reject backgrounds more efficiently and align well with current machine-learning trends in collider physics. Moreover, adopting a hybrid strategy that combines fixed- and dynamic-radius clustering depending on the fat jet's transverse momentum can potentially enhance the reach and robustness of future studies.

ACKNOWLEDGMENTS

The authors sincerely acknowledge the organizers of the WHEPP XVII, 2023, for providing a platform to meet and initiate this work. The authors acknowledge the organizers, instructors, and facilitators of the CMS Open Data Workshop at WHEPP XVII, 2023, for the necessary training to access and process CMS Open Data samples. The authors are thankful to the computing facility at IISER Kolkata for the computational and data storage resources.

DATA AVAILABILITY

Some of the data that support the findings of this article are openly available [130]. All simulated data corresponding to the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: CROSS SECTIONS AND BRANCHING RATIOS FOR DIFFERENT MASS HYPOTHESES

The signal production rates and decay branching ratios for the chosen mass points are compiled in Table IX.

TABLE IX. Cross sections and branching ratios of various decay modes for different top-partner masses. All other parameters are kept the same as in the benchmark point defined in Eq. (7).

M_T (TeV)	$\sigma(pp \rightarrow T\bar{t} + \bar{T}t)$ (fb)	Γ_T (GeV)	$\mathcal{B}(T \rightarrow tZ)$ (%)	$\mathcal{B}(T \rightarrow th)$ (%)	$\mathcal{B}(T \rightarrow tg)$ (%)	$\mathcal{B}(T \rightarrow bW)$ (%)
1.4	139.81	18.48	41.58	9.23	31.68	17.52
1.6	78.24	27.67	41.95	9.09	31.49	17.47
1.8	45.04	39.43	42.24	9.00	31.31	17.45
2.0	26.51	54.09	42.47	8.94	31.13	17.45
2.2	15.96	71.96	42.67	8.90	30.97	17.46
2.4	9.72	93.36	42.84	8.88	30.81	17.47
2.6	6.04	118.58	42.98	8.86	30.67	17.49
2.8	3.82	147.94	43.11	8.85	30.53	17.51
3.0	2.45	181.76	43.22	8.84	30.41	17.53

APPENDIX B: INPUT OBSERVABLES IMPORTANCE FOR MULTIVARIATE ANALYSIS

Figure 12 illustrates the ranking of the input variables according to their impact on the BDT performance for the fixed and dynamic radius jet clustering approaches.

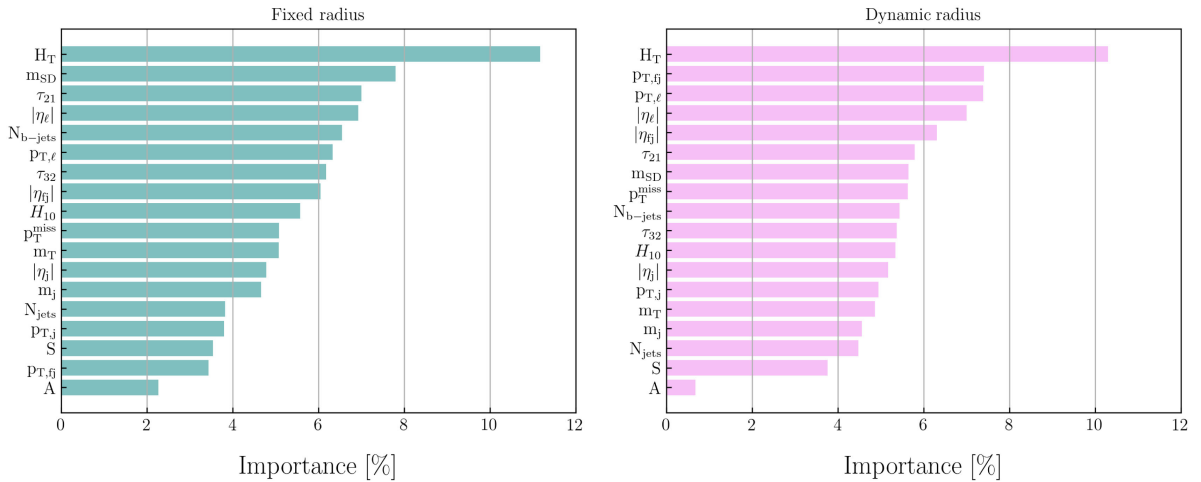


FIG. 12. Importance of input observables to the BDT discriminants corresponding to fixed- (left) and dynamic- (right) radius jet clustering.

- [1] G. F. Giudice, Naturalness after LHC8, *Proc. Sci., EPS-HEP2013* (2013) 163 [arXiv:1307.7879].
- [2] N. Arkani-Hamed, K. Blum, R. T. D’Agnolo, and J. Fan, 2:1 for naturalness at the LHC?, *J. High Energy Phys.* 01 (2013) 149.

- [3] S. Alekhin, A. Djouadi, and S. Moch, The top quark and Higgs boson masses and the stability of the electroweak vacuum, *Phys. Lett. B* **716**, 214 (2012).
- [4] G. Degraasi, S. Di Vita, J. Elias-Miró, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, Higgs mass and

- vacuum stability in the Standard Model at NNLO, *J. High Energy Phys.* **08** (2012) 098.
- [5] S. Dawson and P. Jaiswal, Four generations, Higgs physics, and the MSSM, *Phys. Rev. D* **82**, 073017 (2010).
 - [6] C. D. Froggatt and H. B. Nielsen, Hierarchy of quark masses, Cabibbo angles and CP violation, *Nucl. Phys. B* **147**, 277 (1979).
 - [7] H. Fritzsch and Z.-Z. Xing, Mass and flavor mixing schemes of quarks and leptons, *Prog. Part. Nucl. Phys.* **45**, 1 (2000).
 - [8] A. Riotto and M. Trodden, Recent progress in baryogenesis, *Annu. Rev. Nucl. Part. Sci.* **49**, 35 (1999).
 - [9] M. Dine and A. Kusenko, Origin of the matter-antimatter asymmetry, *Rev. Mod. Phys.* **76**, 1 (2003).
 - [10] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, A direct empirical proof of the existence of dark matter, *Astrophys. J. Lett.* **648**, L109 (2006).
 - [11] Y. Sofue and V. Rubin, Rotation curves of spiral galaxies, *Annu. Rev. Astron. Astrophys.* **39**, 137 (2001).
 - [12] M. Perelstein, M. E. Peskin, and A. Pierce, Top quarks and electroweak symmetry breaking in little Higgs models, *Phys. Rev. D* **69**, 075002 (2004).
 - [13] O. Matsedonskyi, G. Panico, and A. Wulzer, Light top partners for a light composite Higgs, *J. High Energy Phys.* **01** (2013) 164.
 - [14] M. S. Chanowitz, M. A. Furman, and I. Hinchliffe, Weak interactions of ultra heavy fermions (II), *Nucl. Phys. B* **153**, 402 (1979).
 - [15] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, Handbook of vectorlike quarks: Mixing and single production, *Phys. Rev. D* **88**, 094010 (2013).
 - [16] A. Falkowski, D. M. Straub, and A. Vicente, Vector-like leptons: Higgs decays and collider phenomenology, *J. High Energy Phys.* **05** (2014) 092.
 - [17] S. Dawson and E. Furlan, A Higgs conundrum with vector fermions, *Phys. Rev. D* **86**, 015021 (2012).
 - [18] F. del Aguila, L. Ametller, G. L. Kane, and J. Vidal, Vector-like fermion and standard Higgs production at hadron colliders, *Nucl. Phys. B* **334**, 1 (1990).
 - [19] CMS Collaboration, Review of searches for vector-like quarks, vector-like leptons, and heavy neutral leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV at the CMS experiment, *Phys. Rep.* **1115**, 570 (2025).
 - [20] G. Hiller, T. Höhne, D. F. Litim, and T. Steudtner, Portals into Higgs vacuum stability, *Phys. Rev. D* **106**, 115004 (2022).
 - [21] A. Ghosh and P. Konar, Unveiling desert region in inert doublet model assisted by Peccei–Quinn symmetry, *J. High Energy Phys.* **09** (2024) 104.
 - [22] S. Colucci, B. Fuks, F. Giacchino, L. Lopez Honorez, M. H. G. Tytgat, and J. VandeCasteele, Top-philic vectorlike portal to scalar dark matter, *Phys. Rev. D* **98**, 035002 (2018).
 - [23] A. Ghosh, Unveiling a natural multicomponent dark sector: An inert doublet guided by Peccei–Quinn, *Eur. Phys. J. Plus* **140**, 688 (2025).
 - [24] A. Ghosh, P. Konar, and R. Roshan, Top-philic dark matter in a hybrid KSVZ axion framework, *J. High Energy Phys.* **12** (2022) 167.
 - [25] A. Ghosh and P. Konar, Precision prediction of a democratic up-family philic KSVZ axion model at the LHC, *Phys. Dark Universe* **47**, 101746 (2025).
 - [26] A. J. Larkoski, I. Moult, and B. Nachman, Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning, *Phys. Rep.* **841**, 1 (2020).
 - [27] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object Phenomenology*, Vol. 958 (Springer, New York, 2019), 10.1007/978-3-030-15709-8.
 - [28] R. Kogler *et al.*, Jet substructure at the Large Hadron Collider: Experimental review, *Rev. Mod. Phys.* **91**, 045003 (2019).
 - [29] G. P. Salam, Towards jetography, *Eur. Phys. J. C* **67**, 637 (2010).
 - [30] S. Sapeta, QCD and jets at hadron colliders, *Prog. Part. Nucl. Phys.* **89**, 1 (2016).
 - [31] J. M. Duarte, Novel machine learning applications at the LHC, *Proc. Sci., ICHEP2024* (2025) 012 [arXiv:2409.20413].
 - [32] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, *Nature (London)* **560**, 41 (2018).
 - [33] CMS Collaboration, 2020 CMS data preservation, re-use and open access policy, 2020. 10.7483/OPENDATA.CMS.1BNU.8V1W.
 - [34] ALICE, ATLAS, CMS, and LHCb Collaborations, CERN Open Data Policy for the LHC Experiments, CERN Technical Report CERN-OPEN-2020-013, 2020, 10.17181/CERN.QXNK.8L2G.
 - [35] B. Mukhopadhyaya, T. Samui, and R. K. Singh, Dynamic radius jet clustering algorithm, *J. High Energy Phys.* **04** (2023) 019.
 - [36] G. F. Sterman and S. Weinberg, Jets from quantum chromodynamics, *Phys. Rev. Lett.* **39**, 1436 (1977).
 - [37] A. J. Larkoski, G. P. Salam, and J. Thaler, Energy correlation functions for jet substructure, *J. High Energy Phys.* **06** (2013) 108.
 - [38] D. Kar, W. Nzuza, and S. Sinha, 2B or not 2B, a study of bottom-quark-philic semi-visible jets, *SciPost Phys. Core* **7**, 071 (2024).
 - [39] B. Mukhopadhyaya, S. Samanta, T. Samui, and R. K. Singh, Novel signals for the type-X two-Higgs-doublet scenario at the Large Hadron Collider, *Phys. Rev. D* **108**, 075004 (2023).
 - [40] A. Ghosh, P. Konar, T. Samui, and R. K. Singh, Jet substructure probe on scalar leptoquark models via top polarization, *J. High Energy Phys.* **07** (2025) 145.
 - [41] M. Buchkremer, G. Cacciapaglia, A. Deandrea, and L. Panizzi, Model-independent framework for searches of top partners, *Nucl. Phys. B* **876**, 376 (2013).
 - [42] D. Choudhury, K. Deka, and N. Kumar, Looking for a vectorlike B quark at the LHC using jet substructure, *Phys. Rev. D* **104**, 035004 (2021).
 - [43] S. Balaji, Asymmetry in flavour changing electromagnetic transitions of vector-like quarks, *J. High Energy Phys.* **05** (2022) 015.
 - [44] ATLAS Collaboration, Search for new particles in final states with a boosted top quark and missing transverse momentum in

- proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *J. High Energy Phys.* **05** (2024) 263.
- [45] CMS Collaboration, Search for a vector-like quark $T' \rightarrow tH$ via the diphoton decay mode of the Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV, *J. High Energy Phys.* **09** (2023) 057.
- [46] CMS Collaboration, Search for production of a single vectorlike quark decaying to tH or tZ in the all-hadronic final state in pp collisions at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **110**, 072012 (2024).
- [47] A. Hayreter and G. Valencia, T-odd correlations from the top-quark chromoelectric dipole moment in lepton plus jets top-pair events, *Phys. Rev. D* **93**, 014020 (2016).
- [48] CMS Collaboration, Search for CP violation using $t\bar{t}$ events in the lepton + jets channel in pp collisions at $\sqrt{s} = 13$ TeV, *J. High Energy Phys.* **06** (2023) 081.
- [49] CMS Collaboration, Search for pair production of vectorlike quarks in leptonic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, *J. High Energy Phys.* **07** (2023) 020.
- [50] CMS Collaboration, Search for single production of a vector-like T quark decaying to a top quark and a neutral scalar boson in lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV, [arXiv:2510.25874](https://arxiv.org/abs/2510.25874).
- [51] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, *J. High Energy Phys.* **05** (2014) 146.
- [52] J. Thaler and K. Van Tilburg, Identifying boosted objects with N -subjettiness, *J. High Energy Phys.* **03** (2011) 015.
- [53] J. D. Bjorken and S. J. Brodsky, Statistical model for electron-positron annihilation into hadrons, *Phys. Rev. D* **1**, 1416 (1970).
- [54] J. F. Donoghue, F. E. Low, and S.-Y. Pi, Tensor analysis of hadronic jets in quantum chromodynamics, *Phys. Rev. D* **20**, 2759 (1979).
- [55] ATLAS Collaboration, Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV, *Eur. Phys. J. C* **72**, 2211 (2012).
- [56] G. C. Fox and S. Wolfram, Event shapes in e^+e^- annihilation, *Nucl. Phys.* **B149**, 413 (1979); **B157**, 543(E) (1979).
- [57] C. Bernaciak, M. S. A. Buschmann, A. Butter, and T. Plehn, Fox-Wolfram moments in Higgs physics, *Phys. Rev. D* **87**, 073014 (2013).
- [58] 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.
- [59] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, The automation of next-to-leading order electroweak calculations, *J. High Energy Phys.* **07** (2018) 185.
- [60] C. Bierlich *et al.*, A comprehensive guide to the physics and usage of PYTHIA 8.3, *SciPost Phys. Codebases* **2022**, 8 (2022).
- [61] DELPHES 3 Collaboration, Delphes 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [62] N. D. Christensen and C. Duhr, FeynRules—Feynman rules made easy, *Comput. Phys. Commun.* **180**, 1614 (2009).
- [63] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, FeynRules 2.0—A complete toolbox for tree-level phenomenology, *Comput. Phys. Commun.* **185**, 2250 (2014).
- [64] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, UFO—The Universal FeynRules output, *Comput. Phys. Commun.* **183**, 1201 (2012).
- [65] NNPDF Collaboration, Parton distributions for the LHC Run II, *J. High Energy Phys.* **04** (2015) 040.
- [66] ATLAS Collaboration, Study of Higgs boson pair production in the $HH \rightarrow b\bar{b}\gamma\gamma$ final state with 308 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV and 13.6 TeV by the ATLAS experiment, [arXiv:2507.03495](https://arxiv.org/abs/2507.03495).
- [67] J. Zhang, H.-Y. Jin, and T. G. Steele, One loop corrections to the trilinear self coupling of the Higgs boson in the Standard Model with inclusion of singlet vector-like top quark, [arXiv:2510.14425](https://arxiv.org/abs/2510.14425).
- [68] CMS Collaboration, Event generator tunes obtained from underlying event and multiparton scattering measurements, *Eur. Phys. J. C* **76**, 155 (2016).
- [69] P. Skands, S. Carrazza, and J. Rojo, Tuning PYTHIA 8.1: The Monash 2013 tune, *Eur. Phys. J. C* **74**, 3024 (2014).
- [70] CMS Collaboration, Simulated dataset TTJets Tune-CUETP8M1 13 TeV-madgraphMLM-pythia8 in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.4UC1.XE6W](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.4UC1.XE6W), [10.7483/OPENDATA.CMS.TRGR.ISWK](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.TRGR.ISWK), [10.7483/OPENDATA.CMS.QSCL.T5V6](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.QSCL.T5V6), [10.7483/OPENDATA.CMS.P6DS.TA9G](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.P6DS.TA9G), [10.7483/OPENDATA.CMS.3OFU.QCHW](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.3OFU.QCHW), [10.7483/OPENDATA.CMS.ZLCL.5OLD](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.ZLCL.5OLD), [10.7483/OPENDATA.CMS.NIW3.SWVK](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.NIW3.SWVK), [10.7483/OPENDATA.CMS.MY34.FGWJ](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.MY34.FGWJ), [10.7483/OPENDATA.CMS.EAFA.TRYV](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.EAFA.TRYV), [10.7483/OPENDATA.CMS.T3IH.3GLZ](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.T3IH.3GLZ).
- [71] CMS Collaboration, Simulated dataset tW top and antitop inclusiveDecays 13 TeV-powheg-pythia8 Tune-CUETP8M1 in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.EKZ2.EXSU](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.EKZ2.EXSU), [10.7483/OPENDATA.CMS.JG3K.U1HH](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.JG3K.U1HH).
- [72] CMS Collaboration, Simulated dataset $t\bar{t}Z$ Jets 13 TeV madgraphMLM in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.68LT.V5HE](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.68LT.V5HE).
- [73] CMS Collaboration, Simulated dataset TTWJets Tune-CUETP8M1 13 TeV-amcatnloFXFX-madspin-pythia8 in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.Y9YE.GSCH](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.Y9YE.GSCH), [10.7483/OPENDATA.CMS.2J3N.SAEU](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.2J3N.SAEU).
- [74] CMS Collaboration, Simulated dataset WZJets 13 TeV-pythia8 in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.OT9K.8B6X](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.OT9K.8B6X), [10.7483/OPENDATA.CMS.RBRF.G35V](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.RBRF.G35V), [10.7483/OPENDATA.CMS.B0AH.8063](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.B0AH.8063), [10.7483/OPENDATA.CMS.2IZ1.BFHY](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.2IZ1.BFHY).
- [75] CMS Collaboration, Simulated dataset WWJets 13 TeV-powheg in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.8SBL.SMML](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.8SBL.SMML), [10.7483/OPENDATA.CMS.N1NR.ZKCS](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.N1NR.ZKCS).
- [76] CMS Collaboration, Simulated dataset QCD Tune-CUETP8M1 13 TeV pythia8 in MINIAODSIM format for 2015 collision data, 2021, [10.7483/OPENDATA.CMS.9S1H.GPAZ](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.9S1H.GPAZ), [10.7483/OPENDATA.CMS.AQYT.7JKJ](https://cds.cern.ch/record/2811111/files/10.7483/OPENDATA.CMS.AQYT.7JKJ),

- 10.7483/OPENDATA.CMS.XXOE.WFX4, 10.7483/OPENDATA.CMS.2ZYM.S9ZX, 10.7483/OPENDATA.CMS.WFJA.FPGG, 10.7483/OPENDATA.CMS.9OJV.RZ4U, 10.7483/OPENDATA.CMS.O3DT.38EH, 10.7483/OPENDATA.CMS.5DBS.XCW8, 10.7483/OPENDATA.CMS.3Y72.BWW4, 10.7483/OPENDATA.CMS.SGRV.4ABD, 10.7483/OPENDATA.CMS.WOP0.TFY9, 10.7483/OPENDATA.CMS.N49K.1FTR, 10.7483/OPENDATA.CMS.KN7J.TCST, 10.7483/OPENDATA.CMS.OV3H.MOWO, 10.7483/OPENDATA.CMS.RMU2.D9II.
- [77] R. Frederix and S. Frixione, Merging meets matching in MC@NLO, *J. High Energy Phys.* **12** (2012) 061.
- [78] K. Ehatäht and C. Veelken, Stitching Monte Carlo samples, *Eur. Phys. J. C* **82**, 484 (2022).
- [79] E. Re, Single-top Wt -channel production matched with parton showers using the POWHEG method, *Eur. Phys. J. C* **71**, 1547 (2011).
- [80] P. Nason, A new method for combining NLO QCD with shower Monte Carlo algorithms, *J. High Energy Phys.* **11** (2004) 040.
- [81] S. Frixione, P. Nason, and C. Oleari, Matching NLO QCD computations with parton shower simulations: The POWHEG method, *J. High Energy Phys.* **11** (2007) 070.
- [82] S. Alioli, P. Nason, C. Oleari, and E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX, *J. High Energy Phys.* **06** (2010) 043.
- [83] F. Maltoni, G. Ridolf, and M. Ubiali, b-initiated processes at the LHC: A reappraisal, *J. High Energy Phys.* **07** (2012) 022.
- [84] P. Nason and G. Zanderighi, W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2, *Eur. Phys. J. C* **74**, 2702 (2014).
- [85] M. Mangano, The so-called MLM prescription for ME/PS matching, 2002.
- [86] S. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, A. Schaliche *et al.*, Matching parton showers and matrix elements, [arXiv:hep-ph/0602031](https://arxiv.org/abs/hep-ph/0602031).
- [87] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions, *J. High Energy Phys.* **01** (2007) 013.
- [88] T. Sjöstrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [89] GEANT4 Collaboration, Geant4—A simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect. A* **506A**, 250 (2003).
- [90] M. Czakon and A. Mitov, TOP++: A program for the calculation of the top-pair cross-section at hadron colliders, *Comput. Phys. Commun.* **185**, 2930 (2014).
- [91] N. Kidonakis and N. Yamanaka, Higher-order corrections for tW production at high-energy hadron colliders, *J. High Energy Phys.* **05** (2021) 278.
- [92] A. Kulesza, L. Motyka, D. Schwartzländer, T. Stebel, and V. Theeuwes, Associated production of a top quark pair with a heavy electroweak gauge boson at NLO + NNLL accuracy, *Eur. Phys. J. C* **79**, 249 (2019).
- [93] L. Buonocore, S. Devoto, M. Grazzini, S. Kallweit, J. Mazzitelli, L. Rottoli, and C. Savoini, Precise predictions for the associated production of a W boson with a top-Antitop quark pair at the LHC, *Phys. Rev. Lett.* **131**, 231901 (2023).
- [94] M. Grazzini, S. Kallweit, D. Rathlev, and M. Wiesemann, $W^\pm Z$ production at hadron colliders in NNLO QCD, *Phys. Lett. B* **761**, 179 (2016).
- [95] M. Grazzini, S. Kallweit, S. Pozzorini, D. Rathlev, and M. Wiesemann, W^+W^- production at the LHC: Fiducial cross sections and distributions in NNLO QCD, *J. High Energy Phys.* **08** (2016) 140.
- [96] S. D. Ellis and D. E. Soper, Successive combination jet algorithm for hadron collisions, *Phys. Rev. D* **48**, 3160 (1993).
- [97] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions, *Nucl. Phys.* **B406**, 187 (1993).
- [98] M. Cacciari, G. P. Salam, and G. Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [99] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, Better jet clustering algorithms, *J. High Energy Phys.* **08** (1997) 001.
- [100] M. Wobisch and T. Wengler, Hadronization corrections to jet cross-sections in deep inelastic scattering, in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)* (1998), [arXiv:hep-ph/9907280](https://arxiv.org/abs/hep-ph/9907280).
- [101] D. Krohn, J. Thaler, and L.-T. Wang, Jets with variable R , *J. High Energy Phys.* **06** (2009) 059.
- [102] L. Mackey, B. Nachman, A. Schwartzman, and C. Stansbury, Fuzzy jets, *J. High Energy Phys.* **06** (2016) 010.
- [103] D. Ba, A. S. Dogra, R. Gambhir, A. Tasissa, and J. Thaler, SHAPER: Can you hear the shape of a jet?, *J. High Energy Phys.* **06** (2023) 195.
- [104] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, XCone: N-jettiness as an exclusive cone jet algorithm, *J. High Energy Phys.* **11** (2015) 072.
- [105] S. Mondal, G. Barone, and A. Schmidt, PAIRed jet: A multi-pronged resonance tagging strategy across all Lorentz boosts, *J. High Energy Phys.* **09** (2024) 128.
- [106] J. R. Andersen *et al.*, Les Houches 2017: Physics at TeV colliders Standard Model working group report, [arXiv:1803.07977](https://arxiv.org/abs/1803.07977).
- [107] ATLAS Collaboration, Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *J. High Energy Phys.* **08** (2019) 033.
- [108] 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**, 242001 (2008).
- [109] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, Towards an understanding of jet substructure, *J. High Energy Phys.* **09** (2013) 029.
- [110] A. J. Larkoski and J. Thaler, Aspects of jets at 100 TeV, *Phys. Rev. D* **90**, 034010 (2014).
- [111] A. J. Larkoski, S. Marzani, and J. Thaler, Sudakov safety in perturbative QCD, *Phys. Rev. D* **91**, 111501 (2015).
- [112] CMS Collaboration, Measurement of the Splitting function in pp and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Rev. Lett.* **120**, 142302 (2018).

- [113] J. Thaler and K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness, *J. High Energy Phys.* **02** (2012) 093.
- [114] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [115] D. Bertolini, T. Chan, and J. Thaler, Jet observables without jet algorithms, *J. High Energy Phys.* **04** (2014) 013.
- [116] A. J. Larkoski, D. Neill, and J. Thaler, Jet shapes with the broadening axis, *J. High Energy Phys.* **04** (2014) 017.
- [117] G. P. Salam, E_t^∞ Scheme.
- [118] CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector, *J. Instrum.* **12**, P10003 (2017).
- [119] CMS Collaboration, Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV, *J. Instrum.* **13**, P06015 (2018).
- [120] CMS Collaboration, Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC, *J. Instrum.* **16**, P05014 (2021).
- [121] M. Cacciari and G. P. Salam, Pileup subtraction using jet areas, *Phys. Lett. B* **659**, 119 (2008).
- [122] CMS Collaboration, Pileup mitigation at CMS in 13 TeV data, *J. Instrum.* **15**, P09018 (2020).
- [123] CMS Collaboration, Jet algorithms performance in 13 TeV data, CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017, <https://inspirehep.net/literature/1519093>.
- [124] CMS Collaboration, Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, *J. Instrum.* **13**, P05011 (2018).
- [125] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees* (Chapman and Hall/CRC, London, 2017), [10.1201/9781315139470](https://doi.org/10.1201/9781315139470).
- [126] Y. Freund and R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* **55**, 119 (1997).
- [127] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, TMVA: Toolkit for multivariate data analysis, Proc. Sci., ACAT2007 (**2007**) 040 [[arXiv: physics/0703039](https://arxiv.org/abs/physics/0703039)].
- [128] I. Chakravarti, R. Laha, and J. Roy, *Handbook of Methods of Applied Statistics. Vol. 1, Techniques of Computation* (Wiley, New York, 1967).
- [129] G. Punzi, Sensitivity of searches for new signals, and its optimization, eConf **C030908**, MODT002 (2003), [arXiv: physics/0308063](https://arxiv.org/abs/physics/0308063).
- [130] A. Ghosh, S. Ghosh, S. Mitra, T. Samui, and R. K. Singh, Numerical data for selected figures from “Improving sensitivity of vectorlike top partner searches with jet substructure”, Zenodo, [10.5281/zenodo.18407474](https://doi.org/10.5281/zenodo.18407474) (2026).