DRAFT
CMS Paper

The content of this note is intended for CMS internal use and distribution only

Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13\,\text{TeV}$

The CMS Collaboration

Abstract

A search is presented for massive spin-1 Z' resonances decaying to a top quark and a heavy vector-like top quark partner T. The search is based on a $2.6\,\text{fb}^{-1}$ sample of proton-proton collisions at 13 TeV collected with the CMS detector at the LHC. The analysis is optimized for final states in which the T quark decays to a W boson and a bottom quark. The focus is on all-jet final states in which both the W boson and the top quark decay into quarks that evolve into jets. The decay products of the top quark and of the W boson are assumed to be highly Lorentz-boosted and cannot be reconstructed as separate jets, but are instead reconstructed as merged, wide jets. Techniques for the identification of jet substructure and jet flavour are used to distinguish signal from background events. Several models for Z' bosons decaying to T quarks are excluded at 95% confidence level, with upper limits on the cross section ranging from 0.13 to 10 pb, depending on the chosen hypotheses. This is the first search for a neutral spin-1 heavy resonance decaying to a top quark and a vector-like T quark in the all-hadronic final state.

1 Introduction

Many theoretical models of physics beyond the standard model (SM) predict the existence of heavy bosonic resonances [1–9]. Such resonances include Z$'$ gauge bosons [10–12] and Kaluza–Klein excitations of a gluon in Randall–Sundrum models [13, 14]. In many cases the couplings of these resonances to third-generation SM quarks are enhanced, leading to decay channels containing top quarks.

The CMS and ATLAS Collaborations at the CERN LHC have performed several searches for heavy resonances decaying to top quark-antiquark pairs (t$\bar{\text{t}}$) [15–21], placing very stringent limits on their production cross sections in the accessible kinematic range. However, in models with a heavy gluon [22, 23], a composite Higgs boson [24], or extra spatial dimensions [22, 25], an additional fermionic sector may be present in the form of a nonchiral (or vector-like) fourth generation of quarks. Topologies in which the Z$'$ boson decays into vector-like quarks have not yet been investigated experimentally. This search focuses on the kinematic range in which Z$'$ boson decays to tT dominate over those to TT, where T is a vector-like heavy quark with a charge of two thirds.

Vector-like quarks are fermions whose left- and right-handed components transform in the same way under the electroweak symmetry group of the SM. Consequently, their masses can be generated through direct mass terms in the Lagrangian, rather than via Yukawa couplings. This feature makes theories that include a heavy vector-like quark sector compatible with current Higgs boson measurements [26].

We present results of the first search for neutral spin-1 heavy resonances decaying to a top quark and a vector-like quark, in all-jet final states. The search utilizes data from proton-proton collisions at a centre-of-mass energy of 13 TeV. The analysis is optimized for the T $\rightarrow$ bW decay mode, but also considers the T $\rightarrow$ tH and T $\rightarrow$ TZ decays.

The results of the analysis are compared with the predictions of two theoretical models. The first model [22] is an effective theory with one warped extra dimension that considers only the lowest-energy spin-1 and spin-1/2 resonances to describe the decays of the lightest Kaluza–Klein excitation of the gluon, G*, to one SM particle and one heavy fermion. We consider the specific case where the G* resonance decays to a top quark and a heavy top quark partner T. The model assumes branching fractions ($\mathcal{B}$) to be 50/25/25% for T quark decay to the bW/tH/tZ channels. Benchmark values of $\tan\theta_3 = 0.44$, $\sin\phi_{tR} = 0.6$, and $Y_* = 3$ are used for the model parameters. These benchmark values enhance the decays of the heavy resonance to a SM quark and a vector-like quark. The definitions of the parameters, the choice of their values, and their impact on the cross section are explained in Ref. [22], and the significant discovery potential at the LHC even with a comparatively small integrated luminosity is discussed. This model predicts the existence of other vector-like quarks such as the T$_{5/3}$ quark, with a charge of five-thirds and a mass lower than the mass of the T quark. These other heavy quarks can have a moderate impact on the branching fraction of the G* resonance to Tt, and their contribution is properly taken into account when comparing the model with the results of the analysis. In particular, the T$_{5/3}$ quark becomes relevant when the mass of the G* resonance is twice its mass.

The second model [24] is a minimal composite effective theory of the Higgs boson based on the coset $SO(5)/SO(4)$, describing the phenomenology of heavy vector resonances, with particular focus on their interactions with top quark partners. The results of the analysis are compared with the cross section for the production of a neutral spin-1 resonance ρ_L^0 decaying to a top quark and a heavy top quark partner T. The model assumes T branching fractions to tH/tZ channels of 50/50%. The following are benchmark values of the model parameters: $y_L = c_3 = $

47 $c_2 = 1$, and $g_{\rho_L} = 3$. The model parameters and the choice of benchmark values are described
48 in [24]. This model is used to simulate signal samples.

49 The G* and the ρ_L^0 resonances are candidates for the Z$'$ of this search and are both produced
50 through quark-antiquark pair interactions at the LHC. The kinematic distributions of the decay
51 modes considered are comparable between the two models. Hypothetical top quark flavour-
52 changing neutral currents generated in the interaction between the top quark, Z$'$ boson, and T
53 quark are estimated to be below the reach of current measurements [27] because of the large
54 suppression generated by off-shell effects of the Z$'$ boson and the T quark. The leading order
55 Feynman diagram for the production of the Z$'$ boson and the decay chain under consideration
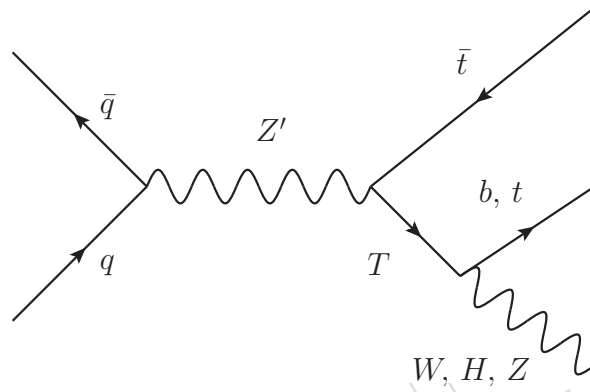56 is depicted in Fig. 1.

Figure 1: The leading order Feynman diagram showing the production mode of the Z$'$ boson
and its decay chain.

57 Because of the large difference in mass between the W boson and the T quark, the W boson
58 receives a large Lorentz boost, such that its decay products appear as merged jets (in a highly-
59 boosted topology). Jet substructure algorithms are employed to reconstruct and identify the W
60 boson originating from the decay of the T quark. If the mass difference between the Z$'$ boson
61 and the T quark is much larger than the mass of the top quark, the top quark from the decay
62 of the Z$'$ boson also receives a large transverse momentum (p_T), in which case jet-substructure
63 techniques can also be used to identify and reconstruct the all-jets decay of the top quark.

64 The dominant background is from SM events and is comprised of jets produced through the
65 strong interaction, i.e. quantum chromodynamics (QCD) multijet events, followed by events
66 from $t\bar{t}$ pair production and from single top quark production. The contribution of the latter
67 processes is estimated from simulation, while the multijet QCD background is estimated from
68 data using signal-depleted control regions.

69 This paper is organized as follows: Section 2 gives a description of the CMS detector and the
70 reconstruction of events. Section 3 describes the data and the simulated samples used in the
71 analysis. An overview of the jet-substructure algorithms and the details of the selection for
72 the analysis are given in Section 4. Estimation of SM background processes is discussed in
73 Section 5, while Section 6 describes the systematic uncertainties. The results of the analysis and
74 a summary are given in Sections 7 and 8, respectively.

75 # 2 The CMS detector

76 The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diame-
77 ter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip

78 tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintilla-
79 tor hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward
80 calorimeters extend the pseudorapidity (η) coverage provided by the barrel and endcap detec-
81 tors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke
82 outside the solenoid.

83 A particle-flow event algorithm [28, 29] reconstructs and identifies each individual particle
84 with an optimized combination of information from the various elements of the CMS detector.
85 The energy of photons is directly obtained from the ECAL measurement, corrected for zero-
86 suppression effects. The energy of electrons is determined from a combination of the electron
87 momentum at the primary interaction vertex as determined by the tracker, the energy of the
88 corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially com-
89 patible with originating from the electron track. The energy of muons is obtained from the
90 curvature of the corresponding track. The energy of charged hadrons is determined from a
91 combination of their momentum measured in the tracker and the matching ECAL and HCAL
92 energy deposits, corrected for zero-suppression effects and for the response function of the
93 calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the
94 corresponding corrected sum of ECAL and HCAL energies. Primary vertices are reconstructed
95 using a deterministic annealing filter algorithm [30]. The vertex with the largest sum of the
96 squares of the associated track p_T values is taken to be the primary event vertex. A more de-
97 tailed description of the CMS detector, together with a definition of the coordinate system used
98 and the relevant kinematic variables, can be found in Ref. [31].

3 Data and simulation samples

100 The analysis is based on data from proton-proton collisions collected in 2015 by the CMS ex-
101 periment at a centre-of-mass energy of 13 TeV, corresponding to a total integrated luminosity
102 of 2.6 fb^{-1}. The events are selected with an online trigger that required the scalar p_T sum of
103 the jets (H_T) to be larger than 800 GeV. The offline H_T is required to be larger than 850 GeV.
104 After this selection, the trigger is more than 97% efficient in selecting those events that would
105 pass the analysis selection. The trigger and offline H_T selections do not significantly impact the
106 overall signal efficiency because the masses of the spin-1 resonances considered in this analysis
107 are at least 1.5 TeV.

108 The signal processes are simulated using MADGRAPH v5.2.2.2 [32]. Neutral spin-1 resonances
109 (Z$'$ boson) decaying exclusively to a top quark and an up-type heavy vector-like quark (T) are
110 generated. Data samples are produced for three values of mass of the Z$'$ boson and a width of
111 1% the mass. For the T quark samples the width of the quark is fixed to 1 MeV. The values of
112 the width are chosen to be much smaller than the detector resolution. The T quark is generated
113 with left-handed chirality. The impact of the chirality of the T quark on the analysis is assessed
114 on a single signal configuration and is found to be insignificant, and for this reason the right-
115 handed chirality case is not explicitly considered.

116 The simulation of the signal event production is based on a simplified low-energy effective
117 theory describing the phenomenology of heavy vector resonances in the minimal composite
118 Higgs model [24]. Signal samples are generated for three decay modes of the T quark: T $\to$ bW,
119 tH, and tZ. Several mass hypotheses for the Z$'$ (T) resonance are considered ranging from 1.5 to
120 2.5 (0.7 to 1.5) TeV. The combination of the Z$'$ and T masses is chosen such that the mass of the
121 T quark is roughly 1/2, 2/3, or 5/6 of the Z$'$ boson mass. For some of the samples generated,
122 the top quark from the decay of the Z$'$ boson receives a small p_T and its decay does not result

in a boosted topology.

The decay of heavy resonances in signal events is processed with MADSPIN [33] to correctly treat the spin correlations in the decay chain. The matrix element calculations for signal processes include one extra parton at most emitted at tree level. To model fragmentation and parton showering, the PYTHIA 8.2 [34] tune CUETP8M1 [35] is used, and the MLM scheme [36] is used to match parton emission in the matrix element with the parton shower. Differential jet rates are checked for smoothness to ensure that the matching scale is chosen correctly.

Background top quark pair production is simulated with the next-to-leading-order generator POWHEG V2 [37–41]. The $t\bar{t}$ event sample is normalized to the next-to-next-to-leading order (NNLO) cross section of $\sigma_{t\bar{t}} = 831.76\,\mathrm{pb}$ [42]. Background events from single top quark production in the tW channel are also generated with POWHEG V2 and are normalized to a cross section of $71.7\,\mathrm{pb}$ [43]. Single top quark production in the s and t channels without an associated W boson is generated with MADGRAPH v5.2.2.2 [32] and the cross sections are normalized to 10.32 and $216.99\,\mathrm{pb}$, respectively [44, 45]. All samples are interfaced to PYTHIA 8.2 for fragmentation and parton showering. The multijet QCD production is estimated from data. Simulated multijet QCD events are used only to validate the method of background estimation and are generated with PYTHIA 8.2, binned in H_T to increase the event sample in the high-energy region.

All events were generated with the NNPDF 3.0 parton distribution functions (PDFs) [46]. All simulated event samples include the simulation of additional inelastic proton-proton interactions within the same or adjacent bunch crossings (pileup). The detector response is simulated with the GEANT4 package [47, 48]. Simulated events are processed through the same software chain as used for collision data and are reweighted to match the observed distribution of the number of pileup interactions in data.

4 Event reconstruction and selection

For each event, hadronic jets are clustered from the reconstructed particles with the infrared and collinear safe anti-k_T algorithm [49], using the FASTJET 3.0 software package [50, 51] with the distance parameters $R = 0.4$ (AK4 jets) and 0.8 (AK8 jets). The two types of jets are reconstructed independently. Charged hadrons not associated with the primary vertex of the interaction are not considered when clustering. Corrections based on the jet area [52] are applied to remove the energy contribution of neutral hadrons arising from pileup collisions. Further corrections are used to account for the nonlinear calorimeter response as a function of η and p_T [53], derived from simulation and from data-to-simulation correction factors. Spurious jets due to detector noise effects are removed by requiring that neutral particles contribute less than 99% of the electromagnetic and hadronic energy in a jet. Only jets with $|\eta| < 2.4$ are considered; no requirements on lepton or imbalance in transverse momentum are applied.

This analysis considers signal events characterized by a three-jet topology. One of the jets corresponds to the boosted top quark from the decay of the Z$'$ boson, the second originates from the W boson of the T quark decay, and the third is from the b quark emitted in the T quark decay. These selection criteria are optimized for the decay of the T quark to bW, but the analysis is sensitive to the other decay modes of the T quark as well. To identify t jets, the jets associated with top quarks, the "CMS top tagger v2" [54] algorithm is used. In this algorithm, the constituents of the AK8 jets are reclustered using the Cambridge–Aachen algorithm [55, 56]. The modified mass-drop tagger algorithm [57], also known as the "soft drop" algorithm with angular exponent $\beta = 0$, soft threshold $z_{\mathrm{cut}} < 0.1$, and characteristic radius $R_0 = 0.8$ [58], is

used to remove soft, wide-angle radiation from the jet. This algorithm identifies two subjets within the AK8 jet corresponding to the b jet and the decay of the W boson. Additionally, the "N-subjettiness" variables τ_N [59, 60] are used. These variables, calculated using all the particle-flow constituents of the AK8 jet, quantify the degree to which a jet can be regarded as composed of N subjets.

For the identification of top quark candidates, the soft-drop mass, m_{SD}, is required to satisfy $110 < m_{SD} < 210\,\text{GeV}$ and the N-subjettiness variable is required to satisfy $\tau_3/\tau_2 < 0.86$. These selections correspond to a misidentification rate of 10% for multijet QCD, and an efficiency greater than 70%. To ensure that the decays of the top quark are merged in a single jet, AK8 jets are required to have $p_T > 400\,\text{GeV}$. Jets satisfying the aforementioned momentum, mass, and N-subjettiness selections are referred to as "t-tagged".

For the identification of W jets, the same jet reclustering procedure as in the t tagging algorithm is chosen. Additionally, jets are required to fulfill $70 < m_{SD} < 100\,\text{GeV}$, $\tau_2/\tau_1 < 0.6$, and $p_T > 200\,\text{GeV}$. These criteria correspond to a misidentification rate of approximately 5% for multijet QCD, and an efficiency of approximately 60% for genuine W bosons not coming from the decay of a top quark. Jets satisfying these requirements are referred to as "W-tagged".

The Combined Secondary Vertex v2 (CSVv2) algorithm [61, 62] is used to identify AK4 jets originating from b quarks (b tagging). The 'medium' working point of the algorithm is used, which provides an efficiency of approximately 70% for the identification of genuine b quark jets while rejecting 99% of light-flavour jets. The 'loose' working point of the algorithm is used for the background estimation, providing an efficiency of approximately 85% and a light-flavour rejection rate of 90%. Additionally, t-tagged jets with a b-tagged subjet [20, 62] are used to improve the discrimination power against background processes. The CSVv2 algorithm with the 'medium' working point is used for subjet b tagging.

The events are required to have at least one b-tagged AK4 jet [61], with $p_T > 100\,\text{GeV}$ and $|\eta| < 2.4$. To avoid possible overlaps, the AK4 jet is required to have an angular separation, ΔR, of at least 0.8 with respect to the t-tagged jet and the W-tagged jet. The angular separation variable ΔR is defined as $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, where ϕ is the azimuthal angle. Among the b jets satisfying these requirements, the one with the highest p_T is selected. The T quark candidate four-momentum is defined as the sum of the 4-vectors of the selected b jet and the W-tagged jet. Only events with a T quark candidate mass $m_T > 500\,\text{GeV}$ are considered. This selection criterion helps to reject the $t\bar{t}$ background. The reconstructed Z' boson candidate four-momentum is defined as the sum of the 4-vectors of the T quark candidate and the selected t-tagged jet. The invariant mass of the Z' boson candidate $m_{Z'}$ is used as the main discriminating observable in the analysis.

Events are grouped into two separate categories according to the presence or absence of a b-tagged subjet in the t-tagged jet. Events containing a b-tagged subjet are placed in the "SR 2 b tag category" as they contain one b-tagged AK4 jet together with a b-tagged subjet associated with the t-tagged jet, as opposed to events in the "SR 1 b tag category" that contain only one b-tagged AK4 jet. No selection criteria are applied to specifically target the tH and tZ final states of the T quark.

Table 1 shows the selection efficiency for the signal in the different event categories. The samples with the smallest difference in mass between the Z' boson and T quark have a degraded reconstruction efficiency because of the low p_T of the top quark originating from the decay of the Z' boson. For several mass points the reconstruction efficiency is higher for the $T \rightarrow tH$ or $T \rightarrow tZ$ decay channel than for $T \rightarrow bW$, for which the analysis is optimized. This is because if

the T quark decays to a t quark instead of a b quark, there are two t quarks in the final state, hence it is more likely that at least one of the two t quarks will be tagged. In addition to this, t quarks coming from the decay of a T quark have a higher p_T, therefore are more likely to be tagged.

Table 1: Selection efficiencies for the signal in the categories used in the analysis. The quoted uncertainties are statistical.

$\mathcal{B}(\mathrm{T} \to \mathrm{bW}) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	1.2 ± 0.2	1.9 ± 0.3
1500	900	0.74 ± 0.17	1.1 ± 0.2
1500	1200	0.23 ± 0.09	0.21 ± 0.09
2000	900	2.6 ± 0.3	3.6 ± 0.4
2000	1200	2.1 ± 0.3	3.0 ± 0.4
2000	1500	0.89 ± 0.18	0.87 ± 0.18
2500	1200	3.3 ± 0.4	3.9 ± 0.4
2500	1500	2.8 ± 0.3	3.6 ± 0.4
$\mathcal{B}(\mathrm{T} \to \mathrm{tH}) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	0.55 ± 0.15	0.75 ± 0.17
1500	900	0.65 ± 0.16	0.93 ± 0.19
1500	1200	0.26 ± 0.10	0.37 ± 0.12
2000	900	1.8 ± 0.3	2.6 ± 0.3
2000	1200	2.0 ± 0.3	2.9 ± 0.3
2000	1500	1.7 ± 0.3	2.2 ± 0.3
2500	1200	2.9 ± 0.3	3.9 ± 0.4
2500	1500	3.0 ± 0.3	4.1 ± 0.4
$\mathcal{B}(\mathrm{T} \to \mathrm{tZ}) = 1$			
$m_{Z'}$ [GeV]	m_T [GeV]	Efficiency SR 1 b tag [%]	Efficiency SR 2 b tag [%]
1500	700	0.62 ± 0.15	0.84 ± 0.18
1500	900	0.78 ± 0.17	0.98 ± 0.19
1500	1200	0.50 ± 0.14	0.54 ± 0.14
2000	900	2.4 ± 0.3	3.1 ± 0.4
2000	1200	2.8 ± 0.3	3.9 ± 0.4
2000	1500	2.3 ± 0.3	2.8 ± 0.3
2500	1200	4.3 ± 0.4	5.4 ± 0.5
2500	1500	4.5 ± 0.4	6.0 ± 0.5

5   Background estimation

There are two dominant source of background: multijet QCD production and top quark production, including both $t\bar{t}$ and single top quark contributions. The multijet background contribution is the most important for this search. Approximately 20% of the top quark production in the signal region is composed of single top quark events, mostly in the tW channel. Pair production of top quarks in association with a vector boson is not a relevant background for this analysis because of the non-boosted nature of the process and its relatively small cross section. Its contribution is estimated to be less than 0.3% of the total number of events in the signal region.

The multijet background is derived from data with the following procedure. Sideband regions are defined by inverting the b tagging requirement on the AK4 jet for the selection of the signal. Specifically, the AK4 jet has to fail the b tagging requirement, using a 'loose' operating point of the b tagging algorithm. Events with additional b-tagged jets are vetoed to ensure independence with respect to the signal region. Two different sideband regions are used for the two signal categories according to the presence or absence of a b-tagged subjet in the t-tagged jet. A summary of the selection criteria is shown in Table 2.

The shape of the $m_{Z'}$ distribution is compared between the sideband region and the signal region in a sample of simulated multijet QCD events. Figure 2 shows the bin-by-bin ratio of the signal region to the sideband region. Both histograms are normalized to unity before computing the ratio.

Table 2: Summary of the selection criteria for the event categories in the signal region (SR) and the sideband region (SB).

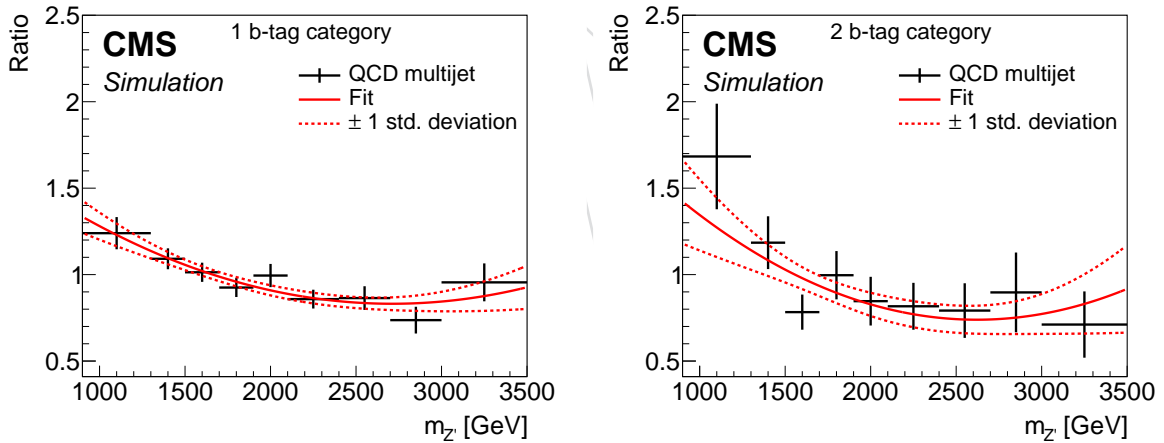| Selection	SR 1 b tag	SB for 1 b tag	SR 2 b tag	SB for 2 b tag
1 t tag and 1 W tag	Yes	Yes	Yes	Yes
Subject b tag on t-tagged jet	Veto	Veto	Yes	Yes
1 AK4 jet, $p_T > 100\,\text{GeV}$, $\Delta R(\text{t} - /\text{W} - \text{jet}, \text{jet}) > 0.8$	Yes	Yes	Yes	Yes
b tag on AK4 jet	Yes	"loose" Veto	Yes	"loose" Veto
$m_T > 500\,\text{GeV}$	Yes	Yes	Yes	Yes

Figure 2: Ratio of the number of events in the signal region to the number in the sideband region, as a function of the Z′ mass, for simulated background QCD multijet events. The left (right) plot involves events with no (at least one) b-tagged subjet. The solid line shows a fit of a second-order polynomial function to the ratio.

The ratio is fit with a second-order polynomial function, which represents the correction factor required to weight the events in the sideband region to reproduce the shape of the multijet background in the signal region. This is the simplest functional form providing a satisfactory fit. To avoid double counting when estimating the multijet background from data, the top quark contribution in the sideband region, estimated from simulation, is subtracted. Good agreement in shape between data and simulated events is observed in the sideband regions.

The normalization of the predicted multijet background cannot be reliably extracted from simulation and is fixed by a maximum likelihood fit to data in the signal region in a background-

²⁴⁶ only hypothesis. The contribution from $t\bar{t}$ is properly taken into account. A flat prior is used
²⁴⁷ for the nuisance parameter associated with the normalization of the multijet background. The
²⁴⁸ fit is performed on the $m_{Z'}$ distribution and, as a consistency check, on the H_T distribution, ob-
²⁴⁹ taining compatible results. It is verified that the scale factor obtained from the fit is not affected
²⁵⁰ by changing the signal hypotheses considered in this analysis. The inclusive normalization
²⁵¹ factors are 0.093 ± 0.004 and 0.12 ± 0.01 for the 1 and 2 b tag event categories, respectively.
²⁵² This normalization is used for plots in Section 7. For the extraction of upper cross section lim-
²⁵³ its on signal production, the normalization of the multijet background is determined by the
²⁵⁴ maximum likelihood fit to data described in Section 7.

²⁵⁵ The top quark background is estimated using simulated event samples normalized to the the-
²⁵⁶ oretical cross sections, as listed in Section 3. The systematic uncertainties that may impact the
²⁵⁷ event rates and the shapes of the $m_{Z'}$ and m_T distributions in simulated events are discussed
²⁵⁸ in Section 6. Table 3 shows the expected background yields for the two event categories, along
²⁵⁹ with the observed number of events in data. The uncertainties include both statistical and sys-
²⁶⁰ tematic components; the estimation of the latter is described in Section 6. The yields have been
²⁶¹ normalized to give the observed total numbers of events.

Table 3: Number of events in the two signal categories of the analysis. The uncertainties include both statistical and systematic components.

Sample	SR 1 b tag	SR 2 b tag
QCD multijet	1227^{+59}_{-59}	222^{+22}_{-22}
SM top quark	81^{+31}_{-23}	66^{+23}_{-18}
Total background	1308^{+67}_{-63}	288^{+32}_{-29}
Data	1307	289

6 Systematic uncertainties

²⁶³ Several sources of systematic uncertainty may impact the simulated signal and the top quark
²⁶⁴ backgrounds. The procedure used to estimate the multijet background is subject to uncertain-
²⁶⁵ ties as well. These systematic uncertainties affect both the shape and the normalization of the
²⁶⁶ $m_{Z'}$ distribution used in the statistical procedure to infer the presence of signal. The systematic
²⁶⁷ uncertainties are treated as nuisance parameters in the likelihood fit used to extract the upper
²⁶⁸ cross section limit on signal production and are constrained by the data. Table 4 reports the
²⁶⁹ sources of systematic uncertainty, their impact on event rates, the type (rate only, or rate and
²⁷⁰ shape), and the processes for which they are relevant.

²⁷¹ The energy scale of jets [53] is corrected with dedicated p_T- and η-dependent factors derived
²⁷² for AK4 and AK8 jets. The jet energy corrections for AK8 subjets are the same as for AK4 jets,
²⁷³ scaled for the difference in jet area. Systematic uncertainties are derived by varying the jet
²⁷⁴ energy scale within its uncertainty and thus obtaining the shape and normalization impact on
²⁷⁵ the distribution of $m_{Z'}$.

²⁷⁶ The energy resolution of jets is lower in data than in simulation, and thus a smearing factor is
²⁷⁷ applied to the four-vectors of AK4 jets, AK8 jets, and to the subjets, in simulated events. The
²⁷⁸ smearing factor for subjets is the same as that for AK4 jets. The impact of this uncertainty,
²⁷⁹ calculated by varying the smearing factor within its uncertainty, is negligible compared to that
²⁸⁰ of the other uncertainties.

Table 4: Sources of systematic uncertainty, their impact on event rates, their type, and the processes for which they are relevant.

Systematic uncertainty	Rate 1 b tag	Rate 2 b tag	Type	Process
b-tagging efficiency	9–14%	12–17%	rate + shape	t bkg and signal
t-tagging efficiency	8–14%	8–14%	rate + shape	t bkg and signal
W-tagging efficiency	0.1–6%	0.1–6%	rate + shape	t bkg and signal
Jet energy scale	0.4–10%	0.1–8%	rate + shape	t bkg and signal
Jet energy resolution	0–2%	0–2%	rate + shape	t bkg and signal
Integrated luminosity	2.3%	2.3%	rate	t bkg and signal
Trigger efficiency	3%	3%	rate	t bkg and signal
PDFs	3–9%	3–8%	rate + shape	t bkg and signal
Pileup reweighting	0–3%	0.1–2 %	rate + shape	t bkg and signal
μ_R, μ_F	3–44%	1–41%	rate + shape	t bkg and signal
Sideband corr. (fit unc.)	4%	9%	rate + shape	QCD multijet
Sideband corr. (fit form)	1%	2%	rate + shape	QCD multijet
Sideband norm.	50%	50%	rate	QCD multijet

The discrepancy of the t tagging efficiency between data and simulation is corrected with scale factors derived in a semileptonic $t\bar{t}$ topology using a "tag-and-probe" technique [63, 64]. This procedure selects a pure sample of $t\bar{t}$ events using a tight selection on the leptonically decaying top quark. The sample is then used to measure the efficiency of the t tagging algorithm on the hadronically decaying top quark. The scale factors are derived as a function of the jet p_T, along with their respective uncertainties. A similar procedure is used to derive the correction factors for the W tagging algorithm. Jet and subjet b tagging efficiency correction factors for heavy- and light-flavour jets [61] are varied within their uncertainties to derive the impact on shape and normalization in simulated samples.

Different choices of the renormalization (μ_R) and factorization (μ_F) scales used to produce the simulated samples induce shape and normalization changes in the Z′ boson mass distribution. The impact is assessed by using dedicated simulated top quark and signal events where the μ_R and μ_F are both scaled up or down by a factor of 2.

The pileup reweighting uncertainty is evaluated by varying the effective inelastic cross section by 5%. To account for trigger efficiency discrepancies in data and simulation, a 3% rate uncertainty is assigned to the simulated signal and top quark event yields. The uncertainty in the measurement of the integrated luminosity is calculated to be 2.3% [65].

The systematic uncertainty related to the choice of the PDF values is assessed by varying the eigenvectors for the NNPDF 3.0 set used in the simulation. The variations are summed in quadrature to obtain the shape and rate variation due to PDF effects.

The systematic uncertainty in the estimation of the multijet background arises from the sideband shape correction function (weight function) as explained in Section 5. When fitting the ratio between the sideband and the signal region, the statistical uncertainties of the simulated samples in the procedure are considered. In addition, a linear functional form for the weight function is tested for comparison, and the observed difference is taken into account as a systematic uncertainty. These uncertainties are propagated through the background estimation procedure to obtain their impact on the shape and normalization of the $m_{Z'}$ distribution. The normalization of the multijet background is determined during the limit setting procedure by allowing it to vary within an uncertainty of 50% in the maximum likelihood fit to data.

The most significant uncertainties are the ones associated with the multijet background fit

function, and with the choice of renormalization and factorization scales. Assigning a 50% uncertainty to the multijet background normalization does not significantly affect the results obtained in Section 7.

7 Results

The $m_{Z'}$ distributions in the two signal categories are shown in Fig. 3. The m_T and H_T variables are shown in Fig. 4. No excess with respect to the expected background is observed.

A template-based shape analysis with the THETA software package [66] is performed, using the $m_{Z'}$ distribution in the two categories, to extract upper cross section limits on a hypothetical signal production. A Bayesian likelihood-based method [27] is used. Expected limit intervals at 95% confidence level (CL) are obtained by performing a large number of pseudo-experiments. The expected background model is varied within the systematic and statistical uncertainties to determine the best fit to the observed data. The modeling of uncertainties in the shapes is performed through cubic-linear template morphing, where the cubic interpolation is used up to one standard deviation and the linear interpolation beyond that [66]. A nuisance parameter is assigned for each systematic uncertainty in the likelihood. For the parameter of interest, i.e. the signal cross section, a uniform prior is used, while log-normal priors are used for the nuisance parameters. The two event categories are fitted simultaneously.

To avoid the normalization of the multijet background being affected by the presence of a hypothetical signal, a prior uncertainty of 50% is assigned in the fit of the signal hypothesis, as discussed in Section 5. The fit is able to constrain the multijet normalization, primarily with the 1 b tag category, which has more events and is less signal enriched.

Table 5 shows the expected and observed limits on the cross section to produce a Z' boson that decays to Tt for different Z' boson and T quark mass hypotheses. Three different hypotheses for the decay of the T quark are considered: 100% branching fraction into bW, tH, or tZ. The effect of increasing the width of the Z' boson or the T quark to 10% on a single signal configuration has been studied and the impact on the cross section limits is found to be negligible in both cases, because of the detector resolution being bigger than 15%.

Figures 5 and 6 show the expected and observed upper cross section limits, respectively, for $Z' \to Tt$ for different hypotheses of the Z' boson and T quark masses, and the branching fraction of the T quark into bW and tH channels, with $\mathcal{B}(T \to tZ) = (1 - \mathcal{B}(T \to bW, tH))$. Observed cross section limits are in all cases within 2 standard deviations of the expected values.

One-dimensional cross section limits compared to the expectation of the composite Higgs boson model [24], as a function of the resonance mass for $m_T = 1.2$ TeV and T branching fraction to tH/tZ channels of 50/50%, are shown in Fig. 7 (left). A comparison of the limits to the warped-extra dimension model [22] for T branching fractions to the bW/tH/tZ channels of 50/25/25% is shown on the right-hand side of the same figure. For some values of the mass of the heavy resonance, the resonance width is predicted to be larger than 10% in the benchmark theoretical models. In these cases the simulated samples do not reproduce the behaviour of the theory benchmarks accurately, hence the cross section values are not considered for the comparison and are marked by a dashed line in Fig. 7. The increase of the total width of the resonance with the increase of its mass is caused by additional decay channels becoming kinematically allowed. The change in slope of the theoretical cross sections around $m_{Z'} = 1.6$ and 2.4 TeV is due to the Tt and TT decay channels respectively becoming kinematically allowed. The comparison with the expectations of theoretical models shows that this search has no sen-
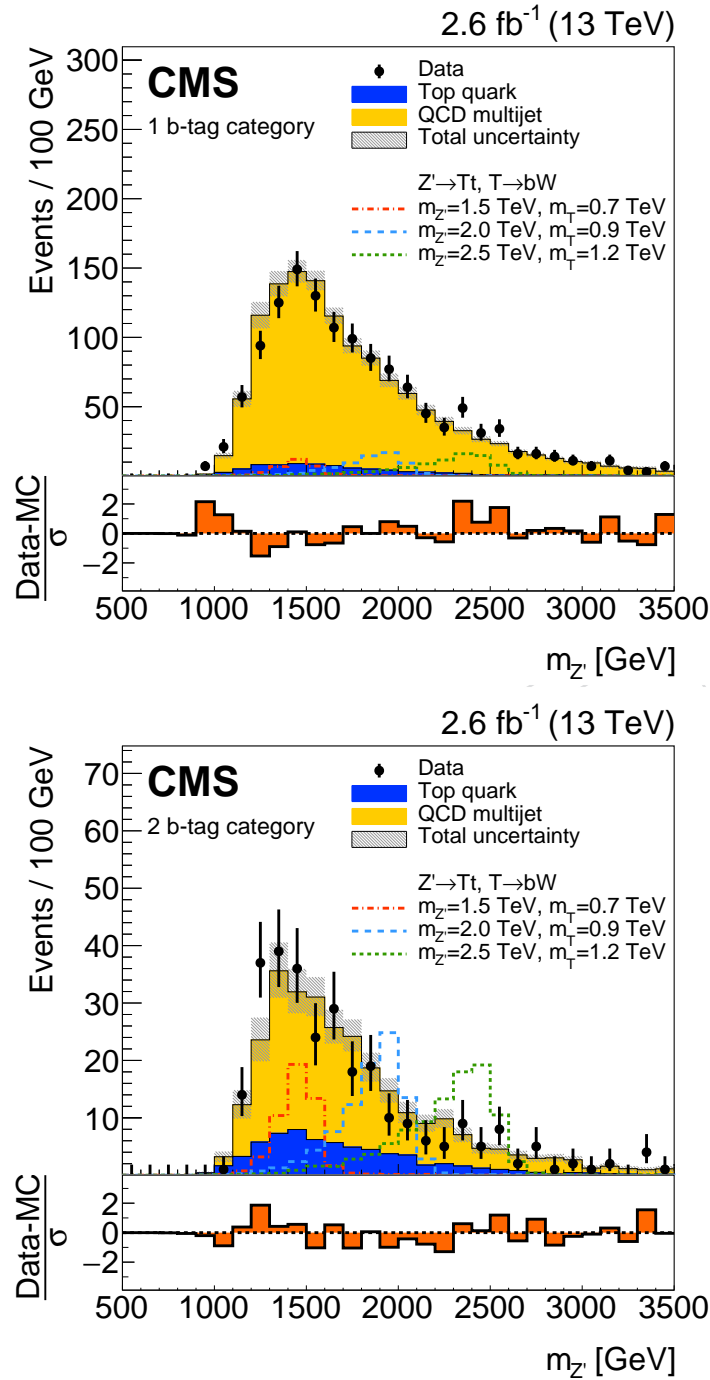
Figure 3: Distribution of the $m_{Z'}$ variable for the signal region with 1 b tag (upper plot) and 2 b tags (lower plot) prior to the fit. The yellow (lighter) distribution represents the multijet background estimated from data, the blue (darker) distribution is the estimated top quark background, and the black markers are the data. The gray bands represent the statistical and systematic uncertainties in the background estimates. The uncertainty σ includes the statistical uncertainties in data and backgrounds, and the systematic uncertainties in the estimated backgrounds. The dashed lines represent the distributions for signal hypotheses as indicated in the legend. The signal distributions are each normalized to a cross section of 1 pb. Events lying outside the x-axis range are not considered.
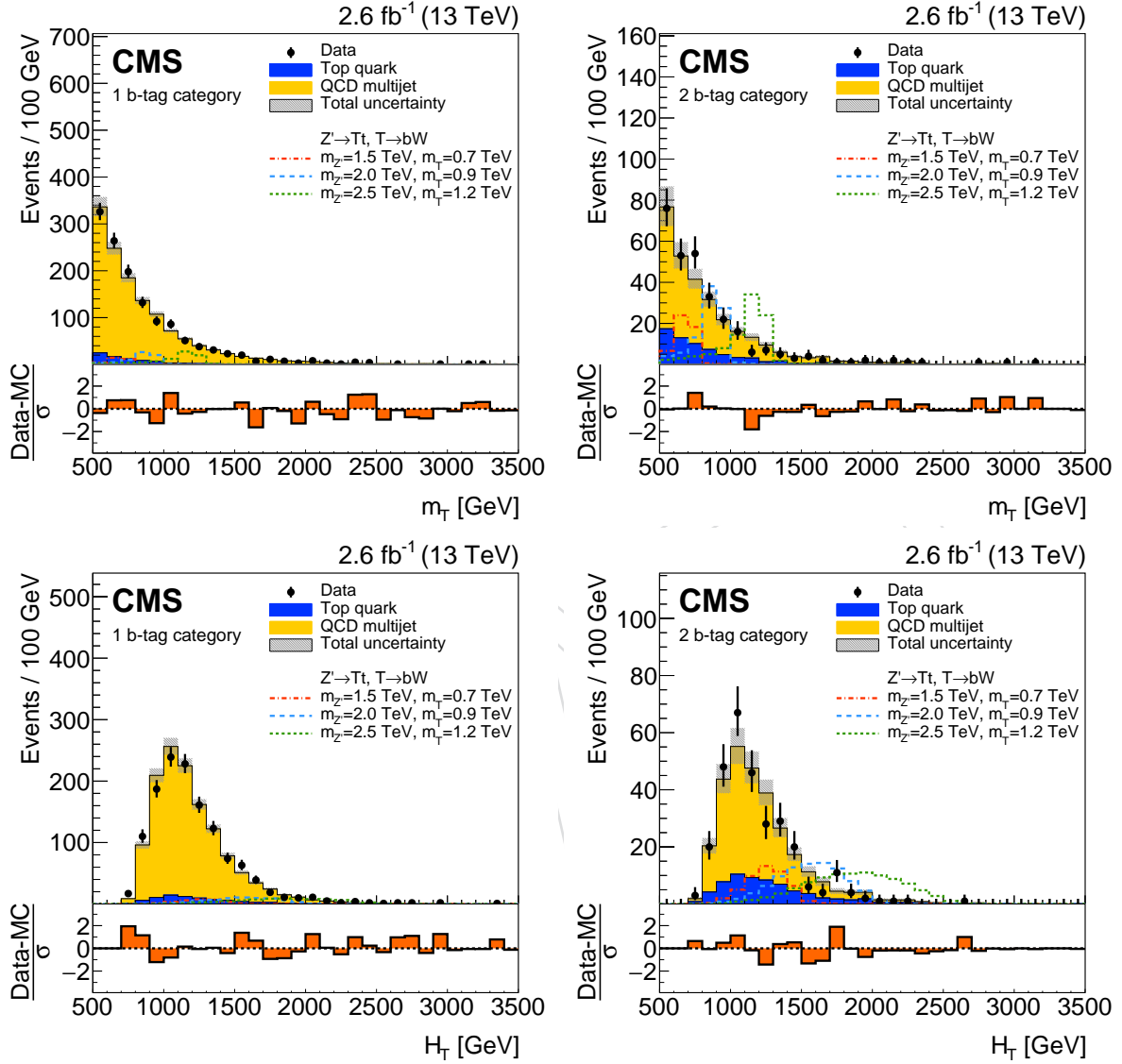
Figure 4: Distributions of the m_T (upper plots) and H_T (lower plots) variables for the 1 b tag (left) and 2 b tag (right) event categories prior to the fit. The gray bands represent the statistical and systematic uncertainties in the background estimates. The uncertainty σ includes the statistical uncertainties in data and backgrounds, and the systematic uncertainties in the estimated backgrounds. The dashed lines represent the distributions for signal hypotheses as indicated in the legend. The signal distributions are each normalized to a cross section of 1 pb. Events lying outside the x-axis range are not considered.

Table 5: Table of expected and observed limits on the cross section to produce a Z' boson that decays to Tt at 95% CL for the $T \to bW$ (upper), $T \to tH$ (middle), and $T \to tZ$ (lower) signal hypotheses.

$\mathcal{B}(T \to bW) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	0.73	0.32	0.48	0.67	1.0	1.6
1500	900	1.5	0.64	0.94	1.5	2.2	3.7
1500	1200	8.6	3.7	5.2	7.8	13	22
2000	900	0.19	0.17	0.24	0.36	0.56	0.90
2000	1200	0.27	0.24	0.33	0.49	0.76	1.3
2000	1500	0.96	0.82	1.2	1.9	3.0	5.4
2500	1200	0.29	0.10	0.15	0.24	0.39	0.64
2500	1500	0.30	0.11	0.16	0.24	0.39	0.65

$\mathcal{B}(T \to tH) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	4.0	0.98	1.4	2.1	3.3	5.8
1500	900	3.2	0.76	1.0	1.6	2.6	4.2
1500	1200	9.4	2.6	3.6	5.6	9.3	19
2000	900	0.53	0.39	0.55	0.84	1.4	2.3
2000	1200	0.53	0.36	0.52	0.79	1.2	2.2
2000	1500	0.60	0.50	0.67	0.99	1.6	2.9
2500	1200	0.24	0.24	0.34	0.52	0.83	1.5
2500	1500	0.23	0.21	0.31	0.49	0.81	1.3

$\mathcal{B}(T \to tZ) = 1$							
$m_{Z'}$ [GeV]	m_T [GeV]	Observed [pb]	Expected [pb]				
			-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
1500	700	3.1	0.84	1.2	1.8	2.9	4.7
1500	900	2.8	0.77	1.1	1.6	2.5	4.3
1500	1200	3.4	1.3	1.8	2.7	4.2	6.4
2000	900	0.37	0.30	0.41	0.61	0.97	1.8
2000	1200	0.30	0.23	0.34	0.50	0.80	1.3
2000	1500	0.32	0.26	0.37	0.55	0.85	1.7
2500	1200	0.16	0.14	0.21	0.31	0.52	0.92
2500	1500	0.13	0.12	0.17	0.27	0.45	0.77
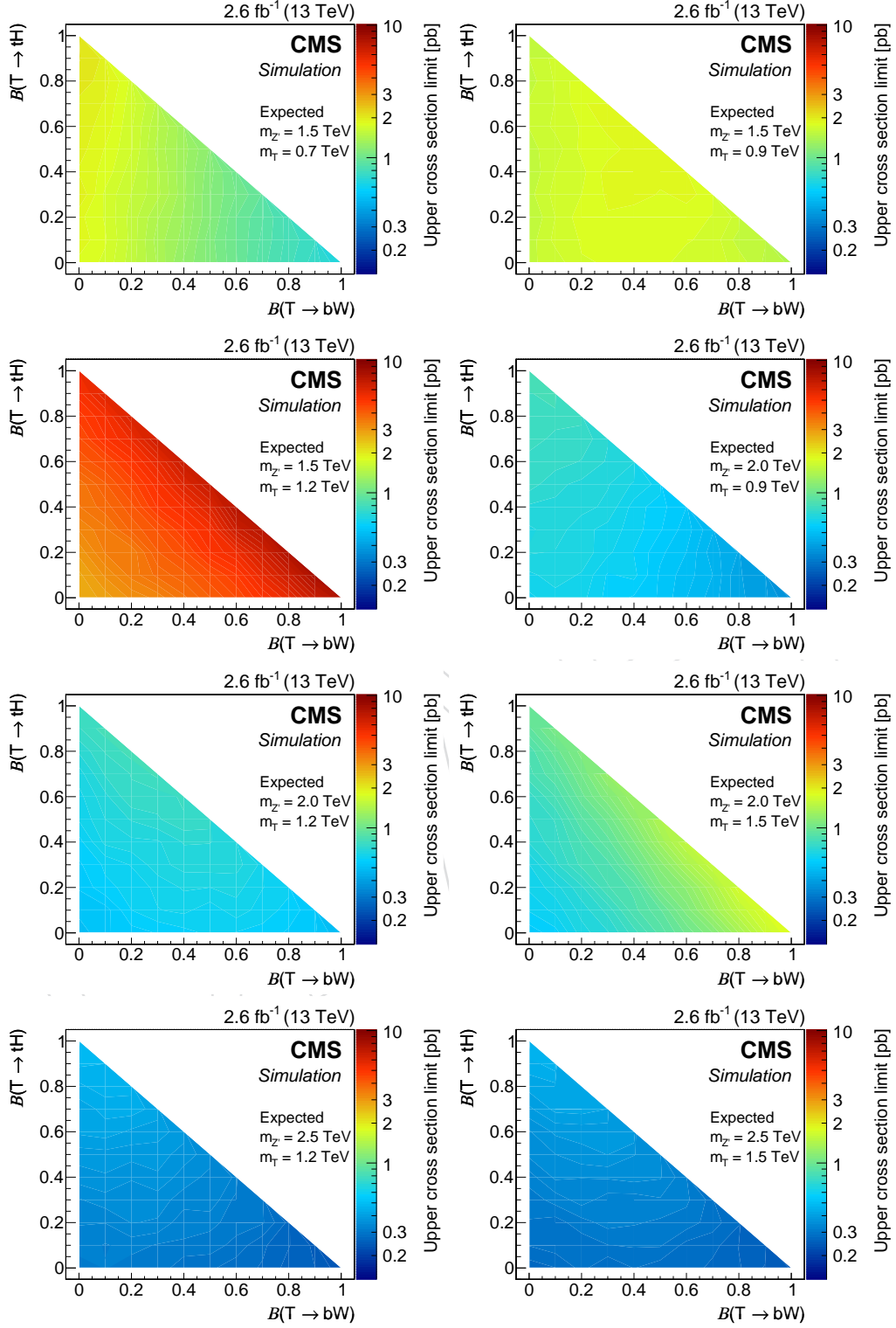
Figure 5: Expected cross section limits for $Z' \to T\bar{t}$ for different hypotheses for the Z' boson and T quark masses, and the branching fraction of the T quark decay into bW and tH channels, with $\mathcal{B}(T \to tZ) = (1 - \mathcal{B}(T \to bW, tH))$.
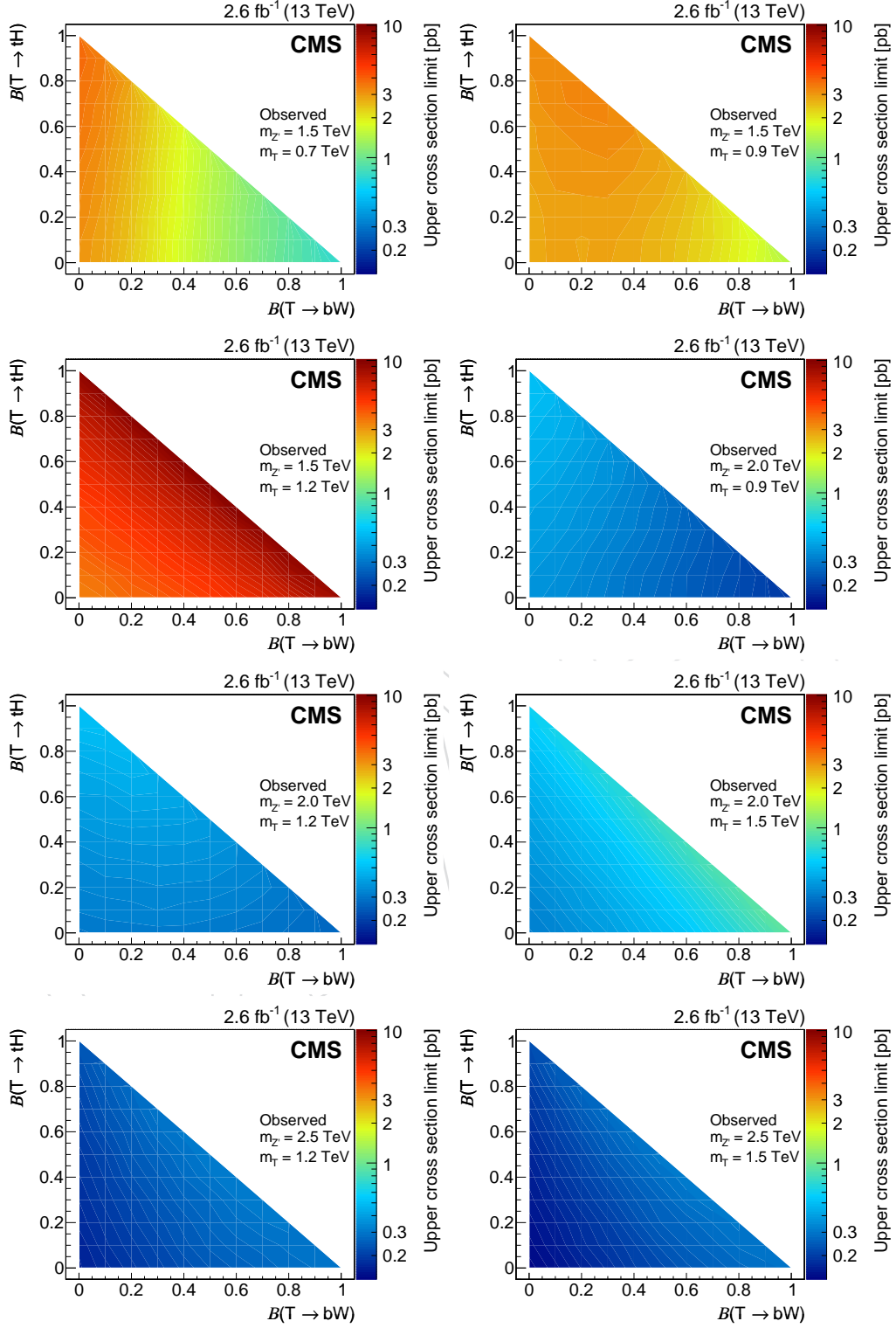
Figure 6: Observed cross section limits for $Z' \to T\bar{t}$ for different hypotheses for the Z' boson and T quark masses, and the branching fraction of the T quark decay into bW and tH channels, with $\mathcal{B}(T \to tZ) = (1 - \mathcal{B}(T \to bW, tH))$.

sitivity to the composite Higgs model [24] and some sensitivity to the extra dimensions model [22], however more data is needed to exclude specific scenarios.
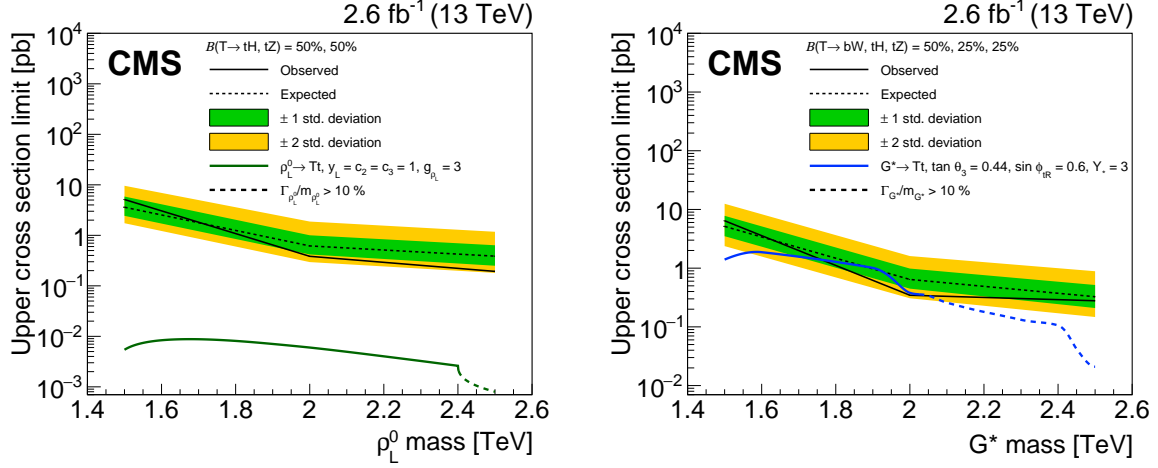
Figure 7: One-dimensional cross section limits at 95% CL as a function of the heavy vector resonance mass for $m_T = 1.2\,\text{TeV}$, assuming branching fractions of the T quark decay to the tH/tZ channels of 50/50% (left) or to the bW/tH/tZ channels of 50/25/25% (right). The solid line is the observed limit, the dotted line is the expected limit, shown with 68% (inner) and 95% (outer) uncertainty bands. In the left plot, the green thick line is the product of the cross section and branching fraction for a heavy spin-1 resonance $\rho_L^0 \to \text{Tt}$ in a composite Higgs boson model [24]. In the right plot, the blue thick line is the product of the cross section and branching fraction for a heavy gluon $G^* \to \text{Tt}$ in a warped extra-dimension model [22]. The theoretical predictions are shown as dashed lines where the width of the resonance is larger than 10% of its mass.

8   Summary

A search for a massive spin-1 resonance decaying to a top quark and a vector-like T quark has been performed in the all-jets channel using $\sqrt{s} = 13\,\text{TeV}$ proton-proton collision data collected by CMS at the LHC. The search uses jet-substructure techniques, involving top quark and W boson tagging algorithms, along with subjet b tagging. The top quark and W boson algorithms are based on the N-subjettiness variables and use the modified mass-drop algorithm to compute the jet mass. The multijet background is estimated in data through a sideband region that is adjusted through simulation-based correction factors. The top quark background is estimated using simulated events.

No excess is observed in data beyond the standard model expectations, and upper limits are set on the production cross sections of hypothetical signals. The cross section limits are compared to the cross sections of a spin-1 resonance in a composite Higgs boson model and a Kaluza-Klein gluon in a warped extra-dimension model, for benchmark values of the model parameters, assuming a T quark mass of 1.2 TeV. Branching fractions of the T quark decay to the tH/tZ channels of 50/50% and to the bW/tH/tZ channels of 50/25/25% are assumed for models with a composite Higgs boson and with a warped extra-dimension, respectively. This search is not sensitive to the composite Higgs model [24] with the analyzed data. In the case of the model with a warped extra-dimension [22], the upper limit obtained on the cross section is just at the predicted level for G^* masses in the region of 1.8 TeV. Although limits are not placed on these particular models, more generally a Z' boson decaying to a top and a T quark is ex-

cluded at 95% confidence level, with upper limits on production cross sections ranging from 0.13 to 10 pb, depending on the hypotheses. This is the first search for a heavy spin-1 resonance decaying to a vector-like T quark and a top quark.

Acknowledgments

References

[1] S. Dimopoulos and H. Georgi, "Softly broken supersymmetry and SU(5)", *Nucl. Phys. B* **193** (1981) 150, `doi:10.1016/0550-3213(81)90522-8`.

[2] S. Weinberg, "Implications of dynamical symmetry breaking", *Phys. Rev. D* **13** (1976) 974, `doi:10.1103/PhysRevD.13.974`.

[3] L. Susskind, "Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory", *Phys. Rev. D* **20** (1979) 2619, `doi:10.1103/PhysRevD.20.2619`.

[4] C. T. Hill and S. J. Parke, "Top production: Sensitivity to new physics", *Phys. Rev. D* **49** (1994) 4454, `doi:10.1103/PhysRevD.49.4454`, `arXiv:hep-ph/9312324`.

[5] R. S. Chivukula, B. A. Dobrescu, H. Georgi, and C. T. Hill, "Top quark seesaw theory of electroweak symmetry breaking", *Phys. Rev. D* **59** (1999) 075003, `doi:10.1103/PhysRevD.59.075003`, `arXiv:hep-ph/9809470`.

[6] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, "Electroweak symmetry breaking from dimensional deconstruction", *Phys. Lett. B* **513** (2001) 232, `doi:10.1016/S0370-2693(01)00741-9`, `arXiv:hep-ph/0105239`.

[7] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The hierarchy problem and new dimensions at a millimeter", *Phys. Lett. B* **429** (1998) 263, `doi:10.1016/S0370-2693(98)00466-3`, `arXiv:hep-ph/9803315`.

[8] L. Randall and R. Sundrum, "A large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370, `doi:10.1103/PhysRevLett.83.3370`, `arXiv:hep-ph/9905221`.

[9] L. Randall and R. Sundrum, "An alternative to compactification", *Phys. Rev. Lett.* **83** (1999) 4690, `doi:10.1103/PhysRevLett.83.4690`, `arXiv:hep-th/9906064`.

[10] J. L. Rosner, "Prominent decay modes of a leptophobic Z'", *Phys. Rev. B* **387** (1996) 113, `doi:10.1016/0370-2693(96)01022-2`, `arXiv:hep-ph/9607207v3`.

[11] C. T. Hill, "Topcolor assisted technicolor", *Phys. Lett. B* **345** (1995) 483, `doi:10.1016/0370-2693(94)01660-5`, `arXiv:hep-ph/9411426`. Updates in `arXiv:hep-ph/9911288`.

[12] K. R. Lynch, M. Narain, E. H. Simmons, and S. Mrenna, "Finding Z' bosons coupled preferentially to the third family at CERN LEP and the Fermilab Tevatron", *Phys. Rev. D* **63** (2001) 035006, `doi:10.1103/PhysRevD.63.035006`, `arXiv:hep-ph/0007286`.

[13] K. Agashe et al., "LHC signals for warped electroweak neutral gauge bosons", *Phys. Rev. D* **76** (2007) 115015, `doi:10.1103/PhysRevD.80.075007`, `arXiv:0810.1497`.

[14] K. Agashe et al., "LHC signals from warped extra dimensions", *Phys. Rev. D* **77** (2008) 015003, `doi:10.1103/PhysRevD.77.015003`, `arXiv:hep-ph/0612015`.

[15] CMS Collaboration, "Search for anomalous $t\bar{t}$ production in the highly-boosted all-hadronic final state", *JHEP* **09** (2012) 29, `doi:10.1007/JHEP09(2012)029`, `arXiv:1204.2488`.

[16] CMS Collaboration, "Search for resonant $t\bar{t}$ production in lepton+jets events in pp collisions at $\sqrt{s} = 7$ tev", *JHEP* **12** (2012) 15, `doi:10.1007/JHEP12(2012)015`, `arXiv:1209.4397`.

[17] ATLAS Collaboration, "Search for resonances decaying into top-quark pairs using fully hadronic decays in pp collisions with atlas at $\sqrt{s} = 7$ tev", *JHEP* **01** (2013) 116, `doi:10.1007/JHEP01(2013)116`, `arXiv:1211.2202`.

[18] ATLAS Collaboration, "A search for $t\bar{t}$ resonances in lepton+jets events with highly-boosed top quarks collected in pp collisions with atlas at $\sqrt{s}$ = 7 tev", *JHEP* **09** (2012) 41, doi:10.1007/JHEP09(2012)041, arXiv:1207.2409.

[19] CMS Collaboration, "Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **111** (2013) 211804, doi:10.1103/PhysRevLett.111.211804, arXiv:1309.2030.

[20] CMS Collaboration, "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s}$ = 8 tev", *Phys. Rev. D* **93** (2016) 012001, doi:10.1103/PhysRevD.93.012001, arXiv:1506.03062.

[21] ATLAS Collaboration, "A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s}$ = 8 tev with the atlas detector", *JHEP* **08** (2015) 148, doi:10.1007/JHEP08(2015)148, arXiv:1505.07018.

[22] C. Bini, R. Contino, and N. Vignaroli, "Heavy-light decay topologies as a new strategy to discover a heavy gluon", *JHEP* **01** (2012) 157, doi:10.1007/JHEP01(2012)157, arXiv:1110.6058.

[23] B. A. Dobrescu, K. Kong, and R. Mahbubani, "Prospects for top-prime quark discovery at the Tevatron", *JHEP* **06** (2009) 001, doi:10.1088/1126-6708/2009/06/001, arXiv:0902.0792.

[24] D. Greco and D. Liu, "Hunting composite vector resonances at the LHC: naturalness facing data", *JHEP* **12** (2014) 126, doi:10.1007/JHEP12(2014)126, arXiv:1410.2883.

[25] N. Vignaroli, "New W' signals at the LHC", *Phys. Rev. D* **89** (2014) 095027, doi:10.1103/PhysRevD.89.095027, arXiv:1404.5558.

[26] ATLAS, CMS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[27] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.

[28] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[29] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.

[30] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[31] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[32] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[33] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[34] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[35] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[36] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[37] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[38] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[39] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[40] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[41] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[42] M. Czakon and A. Mitov, "Top++: a program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[43] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-", *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451. Updates in arXiv:1311.0283.

[44] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[45] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[46] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[47] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[48] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[49] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[50] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[51] M. Cacciari, G. P. Salam, and G. Soyez, "Fastjet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[52] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[53] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[54] CMS Collaboration, "Top tagging with new approaches", CMS Physics Analysis Summary CMS-PAS-JME-15-002, CERN, 2016.

[55] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[56] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep-inelastic scattering", (1998). arXiv:hep-ph/9907280.

[57] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.

[58] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop", *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[59] J. Thaler and K. Van Tilburg, "Identifying boosted objects with N-subjettiness", *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

[60] J. Thaler and K. Van Tilburg, "Maximizing boosted top identification by minimizing N-subjettiness", *JHEP* **02** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.

[61] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, 2016.

[62] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, `doi:10.1088/1748-0221/8/04/P04013`, `arXiv:1211.4462`.

[63] CMS Collaboration, "Boosted top jet tagging at CMS", CMS Physics Analysis Summary CMS-PAS-JME-13-007, CERN, 2014.

[64] CMS Collaboration, "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **01** (2011) 080, `doi:10.1007/JHEP01(2011)080`, `arXiv:1012.2466`.

[65] CMS Collaboration, "CMS luminosity measurement for the 2015 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-15-001, CERN, 2016.

[66] T. Müller, J. Ott, and J. Wagner-Kuhr, "Theta – a framework for template-based modeling and inference", 2010, `http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta.pdf`.