Jet Substructure at the Large Hadron Collider : Experimental Review

Lily Asquith[*][a], Mario Campanelli[b], Chris Delitzsch[†][c], Andreas Hinzmann[d], Deepak Kar[e], Roman Kogler[‡][d], Christine McLean[f], Benjamin Nachman[g], Justin Pilot[f], Alexander Schmidt[§][h], Nhan Tran[i], Caterina Vernieri[i], Marcel Vos[j], and Emma Winkels[a]

[a] *University of Sussex, UK*
[b] *University College London, UK*
[c] *University of Arizona, USA*
[d] *Universität Hamburg, Germany*
[e] *University of Witwatersrand, South Africa*
[f] *University of California, Davis, USA*
[g] *Lawrence Berkeley National Laboratory, USA*
[h] *RWTH Aachen University, Germany*
[i] *Fermilab, USA*
[j] *IFIC Valencia, Spain*

March 22, 2018

Abstract

Jet substructure has emerged to play a central role at the Large Hadron Collider (LHC), where it has provided numerous innovative new ways to search for new physics and to probe the Standard Model, particularly in extreme regions of phase space. In this article we focus on a review of the development and use of state-of-the-art jet substructure techniques by the ATLAS and CMS experiments. ALICE and LHCb have been probing fragmentation functions since the start of the LHC [1, 2] and have also recently started studying other jet substructure techniques [3]. It is likely that in the near future all LHC collaborations will make significant use of jet substructure and grooming techniques. Much of the work in this field in recent years has been galvanized by the Boost Workshop Series [4, 5, 6], which continues to inspire fruitful collaborations between experimentalists and theorists. We hope that this review will prove a useful introduction and reference to experimental aspects of jet substructure at the LHC. A companion overview of recent progress in theory and machine learning approaches is given in [7]; the complete review will be submitted to Reviews of Modern Physics.

Contents

---

[*]lily.asquith@cern.ch
[†]chris.malena.delitzsch@cern.ch
[‡]roman.kogler@uni-hamburg.de
[§]alexander.schmidt@physik.rwth-aachen.de

1   Jet Reconstruction

The majority of particles detected in particle physics experiments are hadrons - bound states of quarks and gluons. Quarks and gluons themselves cannot be observed in isolation due to color confinement, thus observable hadrons are used as a proxy for the quarks and gluons describing perturbative processes. Hadrons are produced in vast quantities by hard-scattering events at the LHC. Most hadrons leave a signal in the calorimeters of CMS and ATLAS (however only those with sufficient energy, roughly 100-500 MeV depending on the noise level, can be detected); charged hadrons additionally leave signals in the tracking detectors.

When hadrons are produced in high energy collisions, the resulting tracks and energy deposits tend to form roughly conical shapes extending from the point of interaction into the detectors. The detectors are arranged cylindrically and symmetrically around the LHC beam. The term 'jet' describes a wide range of objects, from the more coarse objects reconstructed very rapidly in order to identify potentially interesting events for analysis ('trigger-level jets') to the 'physics jets' used at analysis level, which may have gone through a dozen steps of post-processing corrections. Jets have traditionally been used as proxies for quarks and gluons, but they have a rich internal structure that encodes a wealth of information about their origin; the following experimental overview describes techniques for using physics jets as proxies for b quarks,

$W \rightarrow q\bar{q}$ decays, top quarks and others. For a theoretical introduction to jets, we recommend Ref. [8, 9].

1.1   ATLAS and CMS detectors

The ATLAS [10] and CMS [11] detectors are designed to observe leptons, photons, and hadrons resulting from LHC pp collisions. The physics of the hard reaction producing unstable particles and color-confined quarks takes place at the point of collision (the primary vertex) within the beam pipe. Beyond the beam pipe, at 4.4 cm (3.3 cm) in CMS (ATLAS), the first cylindrical layer of detectors encountered are silicon pixels and strips for identification of charged particles. CMS provides a 3.8 T magnetic field via a solenoid positioned outside the silicon tracking detector, the Electromagnetic Calorimeter (ECAL) and most of the Hadronic Calorimeter (HCAL). ATLAS has an additional layer of straw drift tubes (Transition Radiation Tracking : TRT) in the tracking detector, with a 2 T magnetic field encompassing the silicon and TRT, and the ECAL and HCAL situated outside the solenoidal magnet. The calorimeters are surrounded by muon spectrometers which build the outermost part of the ATLAS and CMS detectors.

The energy and momentum ranges and resolutions for the barrel regions[1] of ATLAS and CMS are shown in Table 1 along with the measurement granularity, which limits the angular resolution. The better energy resolution of the CMS ECAL is due to the use of lead tungstate $(PbWO_4)$ crystals, as opposed to the Liquid Argon (LAr) used by ATLAS. The differences in the ATLAS and CMS calorimeter designs are a result of the different ranking of priorities decided by the two collaborations; ATLAS chose a radiation-hard technology with sufficient resolution in a fine sampling LAr calorimeter, while CMS prioritized the excellent resolution of a total absorption crystal calorimeter (the focus was Higgs mass reconstruction), and accepted the accompanying limitations in radiation-hardness associated with this technology. The CMS ECAL crystal response varies under irradiation, which is partially recovered in a few hours at room temperature. The ATLAS ECAL is segmented into three (two) longitudinal layers for $|\eta| < 2.5$ ($|\eta| > 2.5$). The granularity of the ATLAS ECAL in Table 1 refers to its second layer (as most of the electromagnetic energy is deposited there); the first

---

[1]The ATLAS ECAL barrel covers the pseudorapidity range $|\eta| <$ 1.475, the end-caps cover $1.375 < |\eta| < 3.2$ and the forward ECAL layer extends the coverage up to $|\eta| < 4.9$. The CMS ECAL barrel covers $|\eta| < 1.48$, the end-caps extend the coverage up to $|\eta| < 3$.

layer has a finer granularity in η. The multiple layers allow for a finer granularity than the cell size in any of the individual layers, being advantageous over a laterally segmented calorimeter, and additionally provide pointing information. The difference between ATLAS and CMS for the HCAL resolution is particularly large at higher energies: a 1 TeV jet has $\frac{\sigma(E)}{E} \sim 2\%$ in ATLAS, in contrast to $\frac{\sigma(E)}{E} \sim 5\%$ in CMS. To counter this, CMS are using a particle flow approach which includes tracks as discussed in Section 1.2 below.

	ATLAS	CMS
Tracking		
p_T range (GeV)	$p_T > 0.1$	$p_T > 0.2$
$1/p_T$ resolution (%)	$5 \times 10^{-4} p_T \oplus 0.015$	$> 0.8\%$
d_0 resolution (μm)	$140 - 20$	> 9
ECAL		
E resolution (%)	$0.1/\sqrt{E} \oplus 0.007$	$0.03/\sqrt{E} \oplus 0.003$
granularity	0.025×0.025	0.017×0.017
HCAL		
E resolution (%)	$0.5/\sqrt{E} \oplus 0.03$	$1.0/\sqrt{E} \oplus 0.05$
granularity	0.1×0.1	0.087×0.087

Table 1: A comparison of ATLAS and CMS detectors in the barrel regions. The transverse momentum is denoted p_T. The granularity is in pseudorapidity and azimuth ($\eta \times \phi$) for the second layer (inner three layers) of the ATLAS ECAL (HCAL), and d_0 is the transverse impact parameter resolution with respect to the beam-line. Information taken from [11, 12, 13, 14, 15].

1.2 Inputs to Jet Reconstruction

For general jet reconstruction (in events that do not require jet substructure techniques) both ATLAS and CMS use the anti-k_t [16] jet clustering algorithm to form jets from their respective inputs. This *iterative recombination* algorithm takes a single argument: the distance parameter R, which is generally referred to as the jet radius in experimental publications. During the LHC's
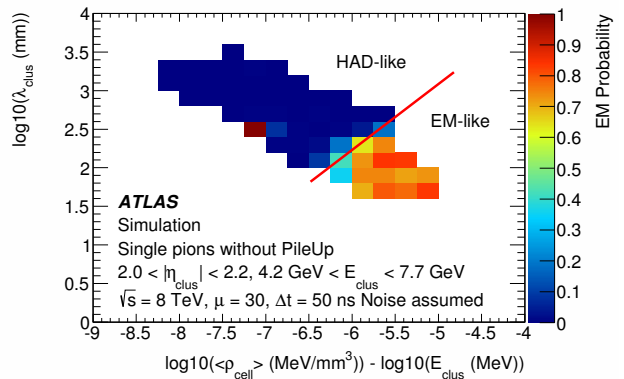

Figure 1: The classification of clusters using the longitudinal depth λ_{clus} and signal density ρ_{cell}. Adapted from [17].

Run 1 (2010-2012), ATLAS favored a distance parameter $R = 0.4$, while CMS favored $R = 0.5$. For Run 2, the two experiments agreed on a common choice of $R = 0.4$ for the sake of better comparability and the additional advantage of the applicability of a theoretical calculation to both experiments.

In ATLAS, three dimensional topological clusters [17] are constructed from the cells of the calorimeter based on their significance, defined as the ratio of the cell signal to the average noise. The topological clustering algorithm implicitly suppresses fluctuations due to electronic and pileup noise (pileup-induced noise dominates since 2011 data-taking) by removing isolated cells with a small cell signal resulting in an improved energy and spatial resolution. The topological clusters are calibrated using the Local Hadronic Cell Weighting (LCW) procedure to account for the non-compensating calorimeter response to hadrons, signal losses due to energy deposited in inactive detector material and signal losses on cluster boundaries caused by the topological clustering algorithm. This calibration scheme relies on classification of clusters as hadronic or electromagnetic in origin based on the energy and position of the cluster, the longitudinal depth (λ_{clus}) and normalized signal energy density, as illustrated in Figure 1; hadronic showers tend to occur deeper in the calorimeter and be less dense.

In recent years, ATLAS has investigated the use of tracks in addition to topoclusters for use in jet reconstruction, the motivation being that the tracker has better transverse momentum (p_T) resolution for low-energy charged particles, and better angular resolution [18]. The *Particle Flow* technique has been shown to improve the
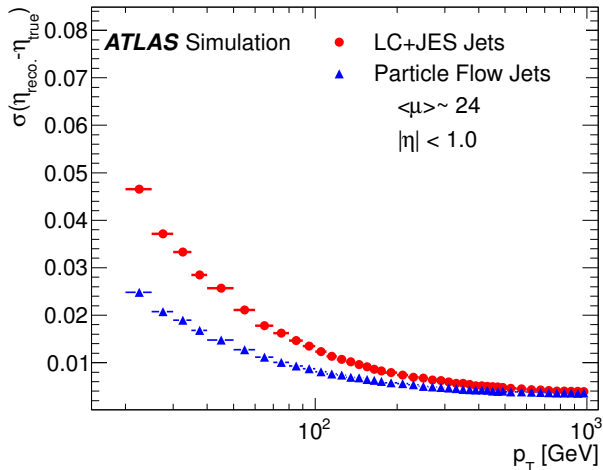
3

Figure 2: Jet η resolution for particle flow (blue) and calorimeter-only (red) jets in the barrel region in ATLAS simulation, as a function of the p_T of the jet. Adapted from [18].

jet energy resolution and reduce the pileup dependence for jets with $p_\text{T} < 100\,\text{GeV}$, by replacing the calorimeter energy deposits from charged hadrons by their measurement in the inner detector [18]. Jets reconstructed with particle flow have improved angular resolution with respect to those reconstructed only from topoclusters, as shown in Figure 2. Tracks are also used by ATLAS for improving the reconstruction of groomed jet masses, as discussed in Section 1.5, and reducing the topology dependence via the Global Sequential Calibration (GSC) (Section 1.3).

Topological clusters from charged-particles inside dense environments (like the core of high p_T jets) can merge, resulting in a loss in tracking efficiency and degraded cluster position resolution. ATLAS has implemented a stacked neural network (NN) approach to examine pixel clusters to identify multi-particle clusters, estimate the position of the particles passing through the clusters, and also predict the residual resolution of the position estimates [19, 20]. This approach significantly improves the efficiency inside jet cores and for boosted hadronically decaying tau leptons. Furthermore, the NN performance has been extensively studied in data and simulation using a variety of methods [20, 21, 22, 23].

In CMS, a particle flow [24, 25, 26] event algorithm is used to reconstruct and identify each individual particle with an optimized combination of information from various elements of the detector. The energy of photons is

directly obtained from the ECAL measurement, corrected for threshold effects in the formation of calorimeter clusters. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for threshold effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy. The momentum for each particle type is calibrated separately based on detector simulation and data measurements. The momentum of charged hadrons is mainly measured with the tracking detector for particle $p_\text{T} < 500\,\text{GeV}$, while at higher p_T the energy measurement of these hadrons comes mainly from the ECAL and HCAL detectors, since the resolution of the tracker for high momentum charged hadrons becomes worse than the ECAL and HCAL resolution. Since the angular resolution of the HCAL is not sufficient to separate close-by particles in jets, the granularity of the ECAL is used by splitting clusters according to their lateral energy profile in the ECAL [27]. The particle flow algorithm dramatically improves the energy response and resolution (Fig. 3) as well as the angular resolution of jets before dedicated corrections, particularly for jets with $p_\text{T} < 100\,\text{GeV}$ as shown in Fig. 4.

For Run 2, CMS has introduced a dedicated tracking step for the dense cores of high-p_T jets. In the innermost layers of the pixel detector charge deposits from separate tracks can be merged into one cluster. A cluster splitting procedure attempts to split such merged clusters exploiting the information of the jet direction, predicting the expected cluster shape and charge. The jet core tracking step also allows a larger number of hit combinations to be evaluated, compared to the standard tracking algorithm. This results in an improvement of the tracking efficiency of up to 10% in the innermost cores of dense jets [28].

1.3 Calibration

The ratio of the measured energy E_{reco} to the deposited energy E_{true} is known as the jet energy *response*, and is
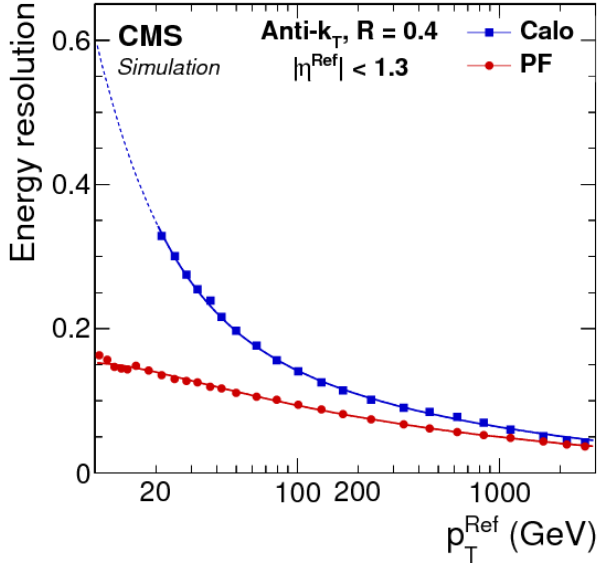
Figure 3: Jet energy resolution for particle flow (red) and calorimeter-only (blue) jets in the barrel region in CMS simulation, with no pileup, as a function of the p_T of the reference jet. Taken from [26].
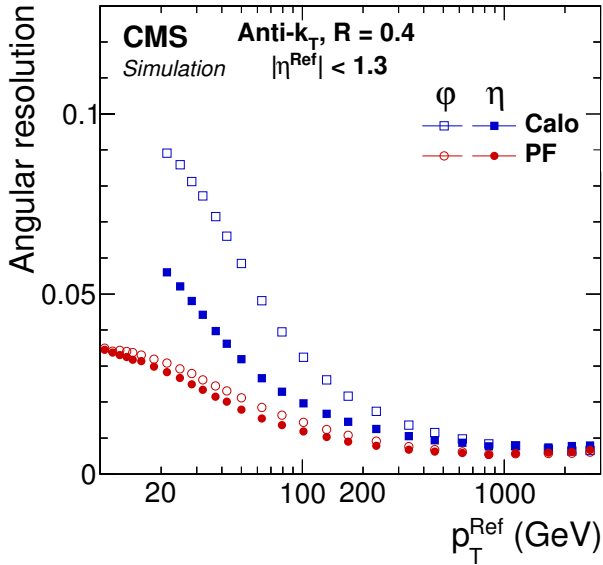

Figure 4: Jet angular resolution for particle flow (red) and calorimeter-only (blue) jets in the barrel region in CMS simulation, with no pileup, as a function of the p_T of the reference jet. The ϕ resolution is expressed in radians. Taken from [26].
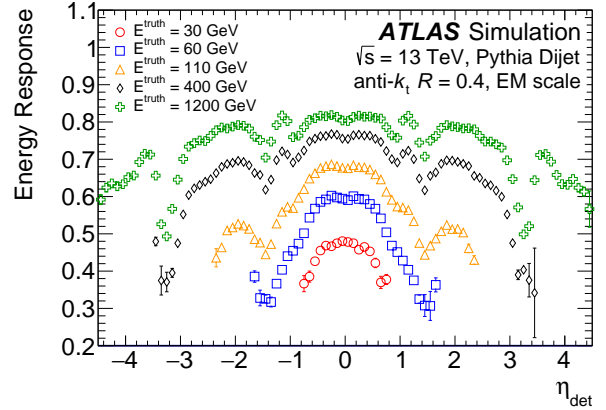

Figure 5: The ATLAS jet energy response prior to calibration, as a function of η for a range of different particle-level jet energies. Adapted from [30].

a crucial aspect of any measurement that uses jets reconstructed from calorimeter energy deposits. The average response is strongly dependent on both the energy and η of the jet, as shown in the pre-calibration distributions of Figure 5. For this reason, the Jet Energy Scale (JES) is calculated in bins of the particle-level jet energy E_{true} and η_{det} as the mean of a Gaussian fit to the response distribution and a numerical inversion procedure is used to derive calibration factors in bins of the reconstructed jet energy from E_{true} [29, 30, 31, 32].

In ATLAS, the calibration of the JES is undertaken in stages, as illustrated in Figure 6 starting from jets either at the electromagnetic (EM) or LCW (built from calibrated inputs) scale. Using calibrated inputs improves the uncalibrated JES by 50% for jets with $E = 30\,\mathrm{GeV}$ and $|\eta| < 0.3$ [29]. The Global Sequential Calibration [30, 33] was introduced for Run 2 and reduces the sensitivity to differences in the responses of quark versus gluon-initiated jets. This additional calibration results in a significant jet p_T resolution improvement of up to 35% depending on the p_T and η of the jet [33]. The JES uncertainty varies between 1-6% in the central region with $\eta = 0$ as shown in Figure 7 [30]. The flavor fraction composition for an unknown quark/gluon composition is of the same order as the absolute *in-situ* uncertainty. The flavor fraction composition is however analysis-dependent and the uncertainty can be reduced if the quark/gluon composition is known. Single-particle response studies and ATLAS combined test beam measurements [34] are used to derive uncertainties for jets with $p_T > 2$ TeV due to the limited statistics of
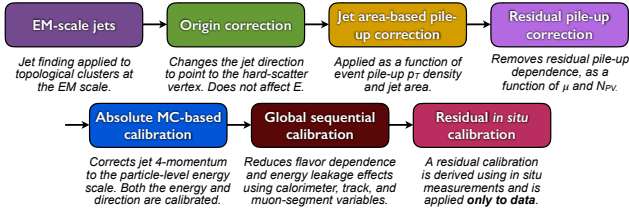
5

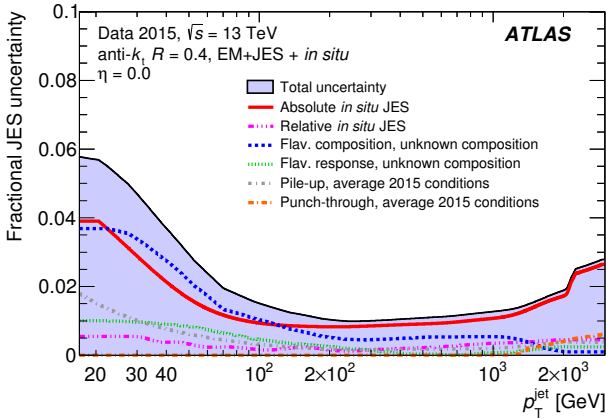Figure 6: ATLAS jet energy scale calibration. Adapted from [30].

Figure 7: ATLAS jet energy scale uncertainty. Adapted from [30].

Figure 8: Simulated jet response versus η for jets with $R = 0.5$ in CMS. Taken from [35].
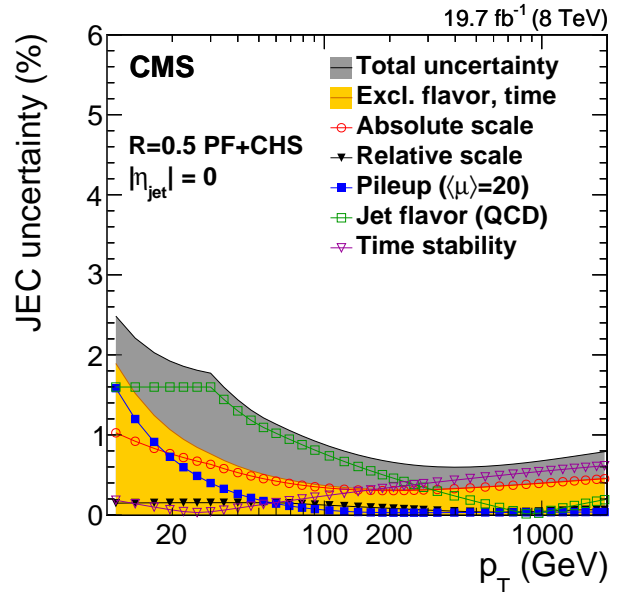
Figure 9: CMS jet energy scale uncertainty, from [35]. JEC means Jet Energy Correction, which has the same meaning as JES.

collision data for the multijet balance technique.

In CMS, jets are clustered from calibrated particles, thus the uncalibrated JES is within 6% of the expected value of 1 for central ($\eta < 0.7$) jets with $p_T > 30$ GeV [35], as shown in Fig. 8. To account for deviations from unity, factorized JES calibrations are applied in multiple stages [36] including pileup corrections, simulation-based response corrections and small residual corrections for tracking inefficiencies and threshold effects, derived *in-situ* from γ+jet, Z+jet and dijet samples [31]. This additional correction is not used when jet substructure observables are constructed, but dedicated corrections are derived as described in Section 1.5. Figure 9 shows the calibrated JES uncertainty obtained in CMS, which is below 1% for jets with $p_T > 100$ GeV in the central region with $\eta = 0$. Even for jet p_T as low as 10 GeV the uncertainty is below 3%, owing to the excellent performance of the particle flow reconstruction.

A detailed discussion of the different approaches for deriving jet energy scale uncertainties in ATLAS and CMS
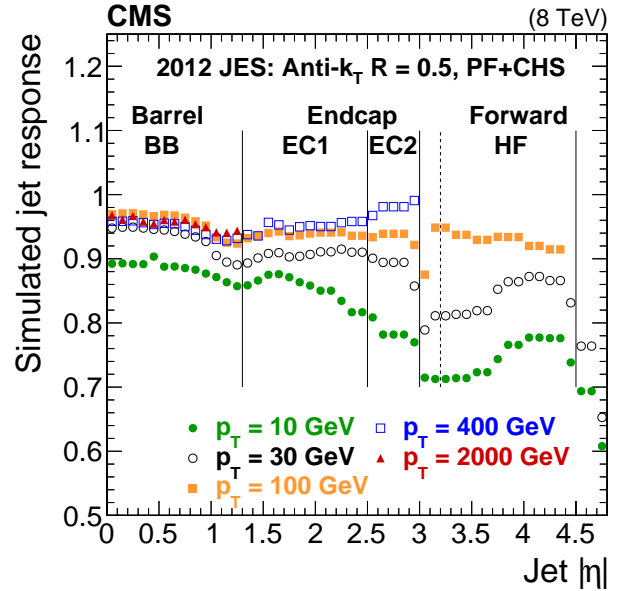
can be found in Ref. [36].

1.4 Jet Grooming

The use of jets as a proxy for quarks and gluons has become an experimental tool of unreckonable importance, particularly at the LHC experiments where jets are present in nearly every event and thus in nearly every analysis. The machinery of jet reconstruction has been evolving at every level over the last decade, from trigger-level to analysis-level, and jet grooming techniques are an area of work that has seen a particularly high level of interest from the experimental and theoretical communities alike. A rule of thumb for the decay of a massive object such as a top quark or $W/Z/H$ boson is that the decay products are likely to lie within a cone of radius $R = 2M/p_T$, where M and p_T are the mass and transverse momentum of the object. Using this rule-of-thumb for the example of a top quark decay, a top quark with $p_T = 350$ GeV will have its decay products captured by a jet of radius $R = 1.0$, and the higher the p_T of the object, the more collimated the decay products. Jet grooming is an additional 'post-processing' treatment of large radius jets, an extra step used to remove unwanted soft radiation and to allow the underlying hard substructure associated with a three-prong (e.g. top quark) or two-prong (e.g. W boson) decay to be utilized in the analysis.

In Run 1, the role of grooming has traditionally satisfied two purposes in ATLAS, being the mitigation of pileup effects on jets and the removal of soft/wide-angle radiation whereas the particle flow algorithm employed in CMS allowed for an efficient correction for pileup effects, reducing the purpose of grooming to the removal of soft/wide-angle radiation. For higher pileup scenarios as currently achieved in Run 2, also the particle flow algorithm is not able anymore to remove all pileup contributions from large-R jets.

ATLAS performed a broad study of the relative performance of different grooming techniques for boson-tagging [37, 38, 39], top-tagging [40, 41] and SM measurements [42, 43], using the removal of pileup-dependence, the jet mass resolution, and the tagging efficiency versus background rejection as performance metrics. The 'standard' grooming procedure adopted by ATLAS is Trimming [44] with $f_{cut} = 0.05$ for boson tagging in both Run 1 (R_{sub} =0.3) and Run 2 (R_{sub} =0.2). The trimming algorithm with the same parameters was adopted for top tagging, along with several other techniques (see Sec. 3.3). An-

other technique currently in use by ATLAS is the *reclustering* of small-R jets [45], which uses fully-calibrated anti-k_t, $R = 0.4$ jets as inputs to the anti-k_t algorithm with a larger distance parameter (typically $R = 1.0$). This has proven a popular method in ATLAS analyses due to the flexibility of optimizing the jet distance parameter depending on the considered phase-space of the analysis [46, 47, 48]. A recent study of *in-situ* measurements [49] (including 'closeby' effects) confirm that the data / MC differences observed with reclustered jets are indeed covered by simply propagating the uncertainties associated with the input anti-k_t, $R = 0.4$ jets.

CMS studied a large number of grooming techniques in the context of boosted boson-tagging [27, 50], top-tagging [51, 52] and SM measurements [53, 54]. During Run 1 the grooming techniques were used together with charged-hadron subtraction for pileup mitigation (see Sec. 2). All groomers studied showed reasonable or good agreement between data and simulation and the Pruning [55] algorithm ($R = 0.8$, $z = 0.1$ and $d_{cut} = 0.5$) showed the best performance for boson tagging [50]. For Run 2 the modified mass-drop tagger (mMDT) [56] is used for jets with $R = 0.8$ in jet substructure analyses in CMS together with the pileup removal algorithm PUPPI [57] (see Sec. 2). In the remainder of this review, jets groomed with the mMDT are referred to as *soft drop* [58] jets, a generalization of the mMDT algorithm, with $z_{cut} = 0.1$ and $\beta = 0$. Soft drop jets in combination with PUPPI show a similar performance as pruning when comparing signal efficiency versus background rejection [52, 59], but allow for better theoretical control. While grooming techniques were found to improve the performance (higher background rejection at fixed signal efficiency) of the jet mass, N-subjettiness ratios [60, 61] were found to perform better without grooming for boosted boson tagging [50]. For top-tagging applications, however, soft drop groomed N-subjettiness ratios improved the performance with respect to ungroomed ones for jets with $p_T < 400$ GeV. For higher p_T jets there was no significant gain observed with grooming for N-subjettiness ratios [52].

1.5 Jet Mass

The reconstruction of jet energies mainly relies on the capability of a detector to measure the total energy of all particles deposited in the detector; however, the measurement of jet mass requires detection of the deposited energy with a granularity that is finer than the size of a jet. The

mass of a jet can only be estimated if the energy is deposited in at least two detector elements, as it depends on both the energy and opening angle between the particles (or detected energy deposits) in a jet. For jet substructure techniques that rely on the rejection of soft particles, it is also important to be able to reconstruct particles with low p_T separately from harder particles in a jet.

The jet mass *response* distribution R_{reco} is constructed from the calibrated, reconstructed jet mass M_{reco} divided by the particle-level jet mass M_{true}. As for the energy response (see 1.3), the mass response distribution is calculated in bins of reconstructed jet $p_{T,reco}$ and η_{reco}. In ATLAS, the Jet Mass Scale (JMS) is defined as the mean of this response distribution. The Jet Mass Resolution (JMR) is then defined as half the 68% interquartile range (IQnR) of the response distribution, as

$$r = 0.5 \times 68\% \text{ IQnR}(R_{reco}). \quad (1)$$

This is robust to large non-Gaussian tails but, if the distribution is Gaussian, is equal to half its 1σ width. The fractional JMR is expressed as the JMR divided by the *median* of the response distribution.

ATLAS has recently developed a data-driven approach to extract the JMS and JMR from an enriched samples of boosted $t\bar{t}$ events, however the method can also be extended to other final states. This *forward-folding* approach folds the particle-level mass spectra by a modified response function such that the average reconstructed jet mass $\langle M_{reco}^{m,p_T} \rangle$ in a given bin of particle-level jet mass and reconstructed jet p_T is scaled by the scale parameter s and the resolution scaled by the resolution parameter r:

$$M_{fold} = s \times M_{reco} + (M_{reco} - \langle M_{reco}^{m,p_T} \rangle)(r-s). \quad (2)$$

The values of r and s for which the M_{fold} distribution best matches the data are extracted from a 2 dimensional χ^2 fit as shown in Figure 10 and detailed in Ref. [62, 63].

With the forward-folding approach, the JMS and JMR for hadronically decaying boosted W bosons with $p_T \gtrsim$ 200 GeV are determined with 2–3% and 20% systematic uncertainties, respectively (see Figure 11). As the jet mass and its detector-response depend on kinematics and jet substructure, the measurement was repeated differentially with an increased luminosity for boosted W and top quarks in Ref. [64]. It will be important to extend the technique (possibly through simulation methods) to other final states in the future. A detailed study of the various
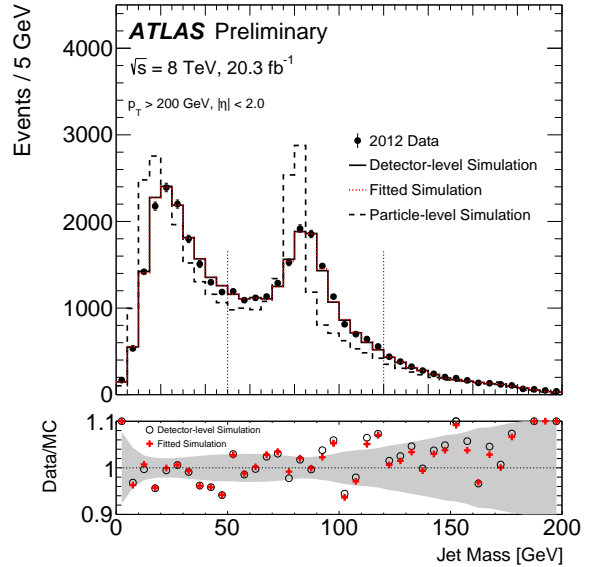

Figure 10: The trimmed jet mass before (detector-level) and after (fitted) determining s and r. The particle-level distribution is shown for comparison. Jets are required to have $p_T > 200$ GeV. Adapted from Ref. [63].

contributions to the JMS and JMR has been performed in context of the soft drop mass measurement [42], described in Section 4.2, by propagating experimental uncertainties on the inputs to the jet reconstruction (topological clusters) to the jet mass. The dominating uncertainties are due to the theoretical modeling of jet fragmentation and the cluster energy scale.

As the forward-folding method is currently restricted to jets with $p_T < 350$ (500) GeV for boosted W bosons (top quarks), the results are combined with the so-called R_{trk} method which constrains the mass scale by comparing the calorimeter jet mass to the mass calculated from track jets and extends up to $p_T = 3000$ GeV [64]. The R_{trk} method can also be generalized to other variables and is used in ATLAS to constrain the p_T scale of large-R jets as well as to derive systematic uncertainties on jet substructure variables.

The concept of a *Track-Assisted Mass* for trimmed, large-R jets has recently been studied in ATLAS [62] to maintain performance for highly boosted particles due to the limited granularity of the calorimeter. The track-
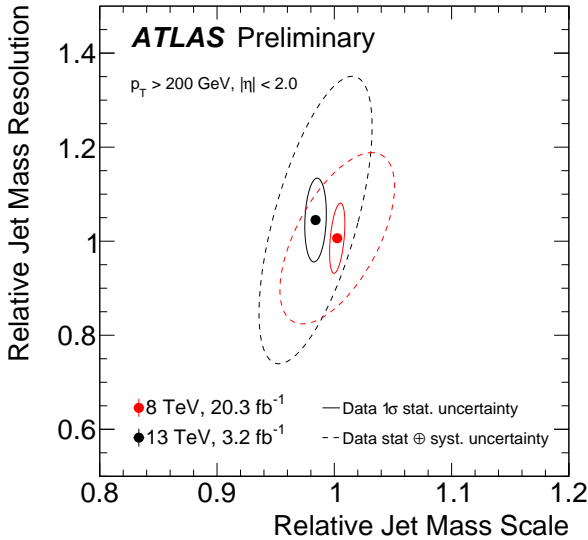
Figure 11: The fitted values of the relative jet mass scale (s) and resolution (r) for trimmed anti-k_t, $R = 1.0$ calorimeter jets from the 2012 and 2015 ATLAS datasets and the 1σ statistical and total uncertainty ellipses. The $\sqrt{s} = 8$ and 13 TeV selections are similar, although the trimming definition slightly changed between Runs (R_{sub} = 0.3 to R_{sub} = 0.2). Adapted from [62].

assisted mass is defined as:

$$m^{\mathrm{TA}} = \frac{p_{\mathrm{T}}^{\mathrm{calo}}}{p_{\mathrm{T}}^{\mathrm{track}}} \times m^{\mathrm{track}}, \qquad (3)$$

where $p_{\mathrm{T}}^{\mathrm{calo}}$ is the transverse momentum of the calorimeter jet, $p_{\mathrm{T}}^{\mathrm{track}}$ is the transverse momentum of the four-vector sum of tracks associated to the calorimeter jet, and m^{track} is the invariant mass of this four-vector sum, where the track mass is set to the pion mass m_π. The track-assisted mass exploits the excellent angular resolution of the tracking detector (whereas the momentum resolution of high-p_{T} tracks is significantly worse than that of the calorimeter) and the ratio $p_{\mathrm{T}}^{\mathrm{calo}}$ to $p_{\mathrm{T}}^{\mathrm{track}}$ corrects for charged-to-neutral fluctuations. The *Combined Mass* is defined as:

$$m^{\mathrm{comb}} = \left(\frac{\sigma_{\mathrm{calo}}^{-2}}{\sigma_{\mathrm{calo}}^{-2} + \sigma_{\mathrm{TA}}^{-2}} \right) m^{\mathrm{calo}} + \left(\frac{\sigma_{\mathrm{TA}}^{-2}}{\sigma_{\mathrm{TA}}^{-2} + \sigma_{\mathrm{calo}}^{-2}} \right) m^{\mathrm{TA}},$$
$$(4)$$

where σ_{calo} and σ_{TA} are the calorimeter-based jet mass resolution and the track-assisted mass resolution, respectively. The jet mass resolution for the calorimeter mass,
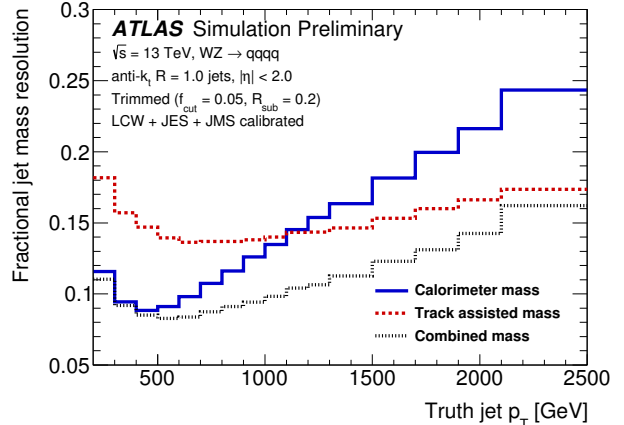

Figure 12: The ATLAS combined jet mass resolution. Adapted from [62].

track-assisted mass and combined mass are shown in Figure 12 as a function of jet p_{T}. Similar techniques that take advantage of the excellent angular resolution of the tracking detector at high p_{T} have been developed to correct topoclusters to improve the resolution of jet substructure variables [65].

It is important to point out that in ATLAS unlike in CMS, the jet energy scale directly impacts the jet mass scale. As opposed to the description of the JES calibration for small-R jets in Section 1.3, the area subtraction, residual correction and GSC calibration are not applied to large-R jets.

In CMS, the jet mass is by default reconstructed as a combination of track and calorimeter measurements via the virtues of the particle flow algorithm. Since the particle flow algorithm reconstructs neutral hadrons from a single energy cluster in the less granular HCAL, which can be associated to multiple clusters in the finer segmented ECAL, the HCAL energy is split according to the number and energy distribution of the ECAL clusters, thereby obtaining an optimal estimate of the spatial energy distribution as input to jet substructure algorithms [27]. Since the residual *in-situ* jet energy corrections are not applied when reconstructing jet masses, dedicated corrections are derived from simulation and data. Firstly, the jet mass response is corrected as a function of p_{T} and η using simulation of W jets from boson pair production. Secondly, residual corrections are obtained from a data sample enriched in lepton+jets $t\bar{t}$ production where the hadronic W jet can be studied in data [50, 59]. The selection is de-
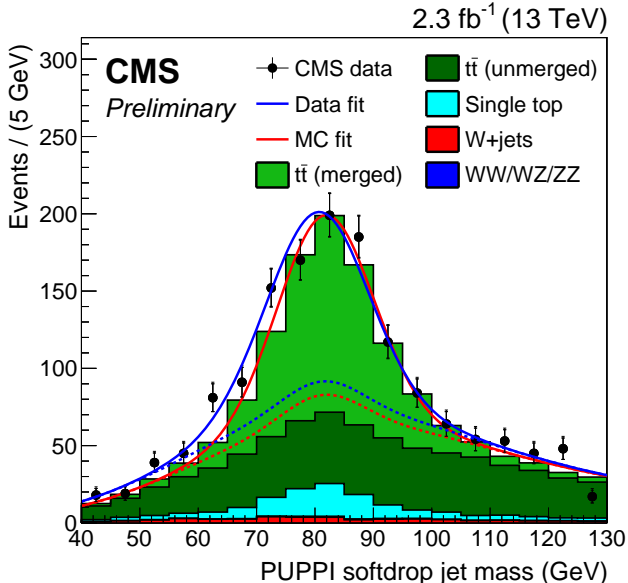
Figure 13: Jet mass distribution in a sample enriched with lepton+jets $t\bar{t}$ events, where the hadronic W jet with $p_T > 200\,\text{GeV}$ is selected, taken from [59].

fined by requiring exactly one isolated lepton (electron or muon), large missing p_T, at least one b-tagged $R = 0.4$ jet, and one groomed (pruning or soft drop) $R = 0.8$ jet with $p_T > 200$ GeV in the final state. The contribution from hadronic W jets in this sample shows a peak at the W mass in the jet mass distribution, as shown in Fig. 13 for the soft drop grooming case. This is used to calibrate the JMS and JMR in CMS, and can also be used for dedicated efficiency corrections on other jet substructure observables, such as the N-subjettiness ratio $\tau_{21} = \tau_2/\tau_1$.

The JMS is described within 1-2% by simulation and the JMR differs by about 10%, which is about the same size as the statistical uncertainty of this measurement. Since these measurements are performed in samples of W jets with $p_T \approx 200$ GeV, additional systematic uncertainties apply at higher p_T [66]. A detailed study of the various contributions to the JMS has also been performed for fully merged top-jets in the context of an unfolded top-jet mass measurement [67]. To summarize the impact of the various sources of systematic uncertainty to the measurement of residual corrections for jet substructure observables, we quote here the dominant uncertainties related to the scale factor measurement of an N-subjettiness ratio $\tau_2/\tau_1 < 0.4$ selection [59]. The statistical uncertainty of

6% (with 2.3/fb of data) is comparable to the systematic uncertainties related to the simulation of the $t\bar{t}$ topology (nearby jets, p_T spectrum) contributing 4%, the choice of method to derive the scale factors contributing 6% and the modeling of the p_T dependence that rises from 5% at $p_T = 500$ GeV to 13% at $p_T = 2000$ GeV.

2 Pileup Mitigation Techniques

2.1 Defining Pileup

Pileup originates from incidental proton-proton (pp) collisions that occur in addition to a hard scattering collision of interest, referred to here as the Primary Vertex (PV). Pileup is uncorrelated with the PV and typically consists of an admixture of inelastic, elastic and diffractive pp processes occurring either *in-time* or *out-of-time* with the hard scatter. Most pileup mitigation techniques focus on in-time pileup, which occurs in the same proton bunch crossing as the PV. The effects of out-of-time pileup (a result of leftover signals from neighboring bunch-crossings) can be substantial, but are often sufficiently suppressed through the detector signal processing.

During the LHC Run 1 the mean number of pileup interactions reached $\langle\mu\rangle = 21$, and μ values up to 60 were achieved in certain runs of 2017 (Run 2) with possibly even higher values in Run 3, and culminating at the high luminosity LHC (HL-LHC) reaching up to $\langle\mu\rangle = 140-200$.

To a good approximation, pileup typically leaves about $0.5\,\text{GeV}$ of energy in the detector per unit area (η, ϕ), per pileup vertex; the effects of this are present in all aspects of LHC physics, from detector design and software performance to the final sensitivity of measurements and searches.

2.2 Mitigation Methods

There are a number of basic properties of pileup interactions which can be exploited in order to discriminate pileup particles from particles originating from the primary vertex. The basic strategies for removing pileup at the LHC are based on these properties, listed below, with examples of some of the commonly-used methods falling in each category.

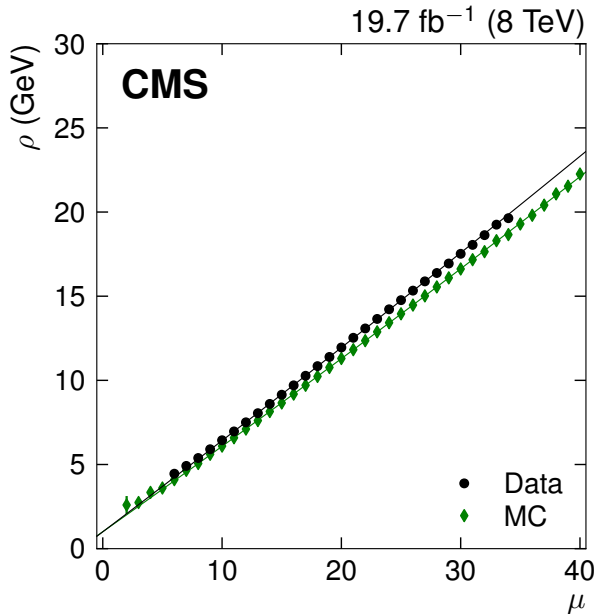(a) Asymptotic behavior: The collective behavior of pileup can be treated as a uniform field of radiation.

Figure 14: Average pileup contribution to the jet p_T, ρ, as a function of the average number of pileup interactions per bunch crossing, μ, for data (circles) and simulation (diamonds) at the CMS experiment. Taken from [35]

Area Subtraction uses a pileup p_T density per unit area estimator, ρ, and defines a jet catchment area, A, to remove energy that originate from pileup interactions from a jet. This corrects the jet naively in the following way: $p_T^{corr} = p_T^{orig} - \rho \times A$. An example of ρ can be seen in Fig. 14. There are many subtleties in defining both ρ and A, which are discussed in e.g. Refs. [68, 69, 70]. Extensions to this method, which corrects only the jet p_T, include Shape Subtraction [71] and Constituent Subtraction [57, 72, 73], the latter of which will be discussed in more detail below.

(b) Charged particle vertexing: Charged particles leave tracks in high granularity tracking detectors at the heart of multi-purpose detectors like ATLAS and CMS and can be separated based on their $\hat{z}$ (along the beamline) position in the beam luminous region (see Fig. 15). Charged Hadron Subtraction (CHS) [69] uses the charged particle tracking information to identify pileup by each track individually. Used in concert with particle flow concepts which attempt to identify each particle in the event uniquely, CHS can effec-

tively remove all charged pileup radiation from the event. Identification of Pileup Jets, formed predominantly from the energy of one or many pileup vertices, is another technique for removing pileup using charged particles; by determining the fraction of energy of the jet from the primary vertex, one can distinguish such pileup jets from the PV jets [70, 74].

(c) Local, topological information: QCD radiation from pileup vertices are often uncorrelated and soft and can thus be removed based on the local energy profile, i.e. if the radiation is not consistent with hard scattering radiation from the PV. This can be done in the transverse plane η, ϕ and also as a function of radiation depth r. Grooming (Sec. 1.4) is a technique to clean the jet of soft and wide-angle radiation which incidentally removes pileup radiation. Topoclustering [17], used by the ATLAS Collaboration, is deployed at the formation of clusters in the calorimeter requiring radiation to have a certain topological profile. In the forward region, where no tracking information is available, jet shapes and topological correlations can be used to identify pileup [75].

(d) Timing information: Detector signals are read out at the LHC bunch frequency (40 MHz) or faster, and can be used to reduce out-of-time pileup. With potential future detectors, precision timing capabilities with $\mathcal{O}(10 \text{ ps})$ could possibly distinguish between pileup vertices within the same pp bunch crossing, as pileup interactions can differ from the PV in time as well as $\hat{z}$ position.

While the above methods have been successfully deployed in the LHC experiments, they each have some deficiencies as well; ideally, one would hope to effectively combine all pileup mitigation handles in order to maximally distinguish pileup from PV radiation and to remove pileup at the most granular level possible, i.e. at the particle or constituent level, in order to be as generic as possible. For example, while area subtraction is very effective for correcting the jet p_T, it is not used to mitigate the pileup dependence of jet substructure observables as it is only able to correctly remove pileup contributions on average. In fact, jet substructure variables are among the most difficult to correct for pileup because they are so reliant on radiation profiles. A number of hybrid methods have been proposed operating at the event constituent level, a few examples of which are Constituent Subtraction [73], Soft-
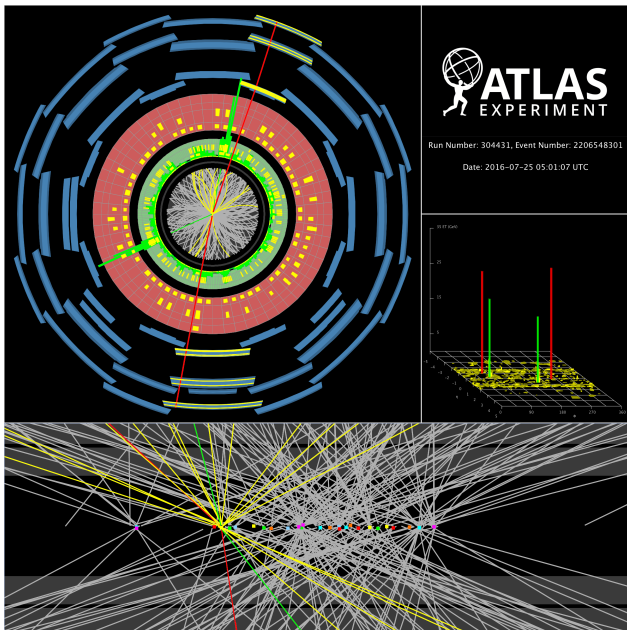
11

Figure 15: $H \to e\mu$ candidate event with 25 additional reconstructed vertices recorded in 2016. Taken from [76].

Killer [72], PUMML [77] and PUPPI [57]. Precursor hybrid methods include *jets without jets* [78] and *jet cleansing* [79]. Most hybrid strategies assume perfect tracking information for identifying charged particles from pileup. Constituent Subtraction and SoftKiller use additional information based on the energy density (ρ) to further remove neutral particles from the event, while PUPPI uses both event energy density and local topological information incorporated in an event-by-event particle-level discriminator to determine if a particle is from pileup. Particle four-vectors are then weighted proportional to the value of the discriminator value.

2.3   Performance Studies

Pileup removal algorithms are commissioned for use in ATLAS and CMS via detailed studies of jet observables in terms of the resolution and absolute scale, pileup dependence, and ultimately the effective background rejection power versus signal efficiency for boosted heavy particle taggers.

For observables like jet p_T, dependencies on the number of reconstructed vertices and μ are observed even with area subtraction methods for the pileup levels currently observed at the LHC, $\langle \mu \rangle \sim 25$. To correct for these ef-
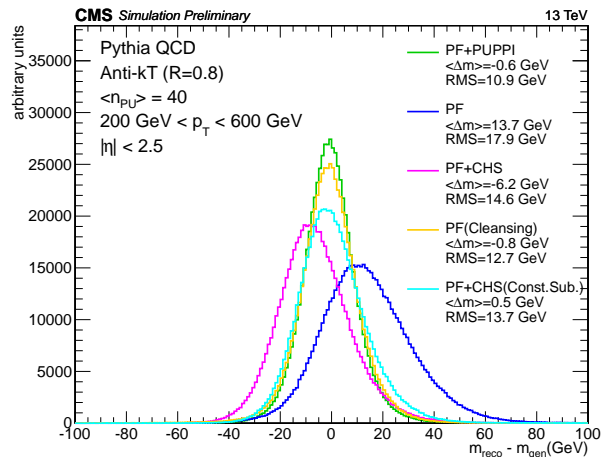

Figure 16: Comparison of different pileup removal algorithms for the leading ungroomed jet mass response in simulated QCD multijet events with CMS. Taken from [69].

fects, an additional residual correction is applied [30, 35]. Enhancements are also possible from combining area subtraction methods with e.g. CHS.

For jet substructure observables, particle- or constituent-level pileup mitigation strategies have been shown to improve performance, especially in simulation studies going up to $\langle \mu \rangle \sim 40$. An example is given in Fig. 16, where the leading ungroomed jet mass in simulated QCD multijet events is corrected with different pileup removal techniques. The jet mass resolution can be improved further when using grooming algorithm. The effect of different pileup removal techniques on the groomed jet mass depends however strongly on the choice of the grooming algorithm as discussed in detail in Ref. [39, 69].

Generally these techniques, particularly those which operate at particle-level, can also be used to improve performance of non-jet objects such as missing transverse energy and lepton isolation.

Preliminary studies (detector configurations have not yet been finalized) into the application of these advanced hybrid techniques at the higher pileup levels anticipated at the HL-LHC suggest that they are effective in the $\langle \mu \rangle = 140 - 200$ range [80, 81].

3 Boosted Particle Tagging

Particle identification is an experimental challenge that is traditionally faced using custom-designed charged-particle detectors, muon chambers and calorimeters with granularity fine enough to allow shower shape measurements. The ATLAS and CMS detectors were designed with particle identification in mind. The use of jet grooming to reveal the jet mass for top quark and boson decays has provided experimentalists with a potential new way to identify high p_T, massive particles - exactly the kind of particles that are most interesting for precision measurements and new physics searches at the LHC. Several substructure variables have been developed by the theoretical community that can be used along with the jet mass for boosted particle identification. The term 'tagger' indicates the use of one or more of these variables (often after grooming has been applied) to discriminate between jets coming from different types of particle. A summary of the tagging algorithms currently in use by ATLAS and CMS is tabulated in Table 2.

3.1 Quark/Gluon Discrimination

In the nearly 40 years since the discovery of the gluon at PETRA [82, 83, 84, 85], jet substructure observables have been widely used for quark-initiated (quark) versus gluon-initiated (gluon) jet tagging. The probability for a gluon to radiate a gluon is enhanced by a factor of $C_A/C_F = 9/4 \sim 2$ over the probability for a quark to radiate a gluon of the same energy fraction and opening angle [86]. As a result, gluon jets tend to have more constituents and a broader radiation pattern than quark jets. There are also more subtle differences due to quark and gluon electric charges and spins.

There are three key challenges of quark versus gluon jet (q/g) tagging: (1) quark and gluon labeling schemes are not unique; (2) for a given labeling scheme, quark and gluon jets are not that different; (3) the differences that do exist are sensitive to both perturbative and non-perturbative modeling choices. Since quarks and gluons carry color charge and only colorless hadrons are observed, there is not a unique way to label a jet in simulation as originating from a quark or a gluon. Many labeling conventions exist, ranging in simplicity and model-dependence from matching to out-going matrix element partons to parsing an entire jet clustering history [87, 88]; however, no treatment escapes the problem that the no-
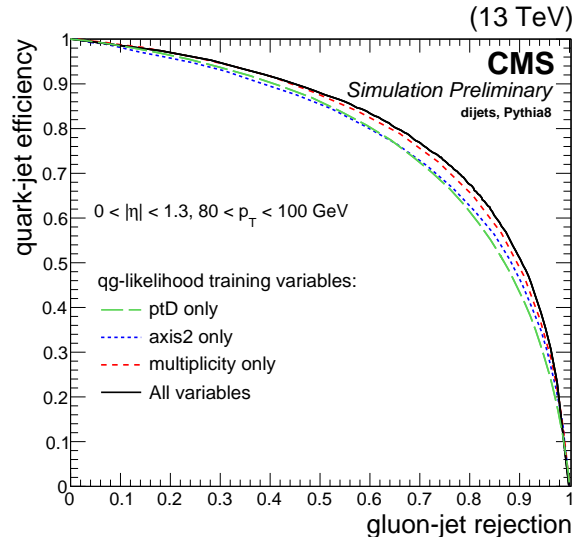

Figure 17: The CMS q/g tagging performance in simulation for a variety of observables. Reproduced from Ref. [59].

tion of a quark and gluon jet are not universal[2]: quark and gluon jet radiation depends on the production mechanism. There is a plethora of jet substructure observables that can be used for q/g tagging; see e.g. Ref. [90] for a large survey. Many of these observables exhibit *Casimir scaling* which results in nearly the same, limited discrimination power for all the observables [91, 92]. The most powerful single q/g observable is the particle multiplicity inside a jet, which does not exhibit Casimir scaling and recent theoretical advances [93] have shown that its discrimination power can be largely understood from perturbative theory. There is further q/g separation possible when using the full radiation pattern inside a jet, though the combination of multiplicity and a Casimir scaling observable carries a significant fraction of the total discrimination power [94]. The modeling of q/g tagging observables has a long history - see Ref. [95] for a recent and detailed study.

Despite the challenges listed above, both ATLAS and CMS extensively use explicit or implicit quark versus gluon tagging. Explicit taggers are algorithms designed to directly isolate quark and gluon jets while implicit techniques are designed for another purpose that also happens to perform some quark versus gluon jet tagging. The explicit taggers developed by ATLAS [96, 97, 98, 99]

---

[2]This can be largely mitigated by jet grooming; see e.g. Ref. [89].

and CMS [59, 100, 101, 102] include a variety of observables and data-driven calibration and validation techniques. These and related techniques have been successfully deployed in a variety of physics analyses (see e.g. [27, 103, 104, 105, 106, 107, 108]). Additionally, it has been shown that an improved W tagger can be constructed by utilizing q/g discrimination on subjets [27].

Both ATLAS and CMS have developed likelihood-based discriminants for explicit q/g tagging. The discriminants are constructed from variables sensitive to the radiation pattern of quark and gluon jets, also taking into account differences between light (uds) and heavy flavor (cb) quark jets, where the latter are more similar to gluon jets. ATLAS uses the number of tracks n_{track} as an approximation for the number of jet constituents and the jet width [99] while CMS utilizes the number of particle-flow constituents n_{const}, the jet axes and fragmentation functions [59]. Since the distributions of these variables depend on η, p_T, and ρ, the likelihood discriminators are constructed differentially with respect to these variables. In Run 2, ATLAS also introduced a simple and robust tagger using solely n_{track} [98], which has the advantage of a much-simplified uncertainty derivation.

Figure 17 shows the CMS q/g tagging performance in simulation. For a 50% gluon (quark) efficiency, the quark (gluon) efficiency is about 10%. This performance depends slightly on the jet p_T, in part because the particle multiplicity increases with p_T (and therefore the performance improves). The p_T and $|\eta|$ dependence is illustrated for the same tagger in Fig. 18. Outside the tracking acceptance ($|\eta| \gtrsim 2.5$), q/g tagging significantly degrades due to the coarse calorimeter granularity and increased pileup sensitivity.

ATLAS [99] and CMS [102] are also actively studying sophisticated approaches based on modern machine learning. While these methods hold great promise for their power and flexibility, Fig. 19 indicates that the deep convolutional neural network (CNN) has similar performance to a 2D binned likelihood ratio (LLH) based on only two variables for lower efficiencies whereas promising performance gains are observed for high efficiencies. Architecture design and input optimization are still an active area of research and development.

The modeling of q/g discriminating observables is a key concern for tagging applications. Typically, Pythia tends to describe quarks better than Herwig, whereas the opposite is observed for gluons. Figure 20 illustrates the typical magnitude of the mis-modeling (the differences between
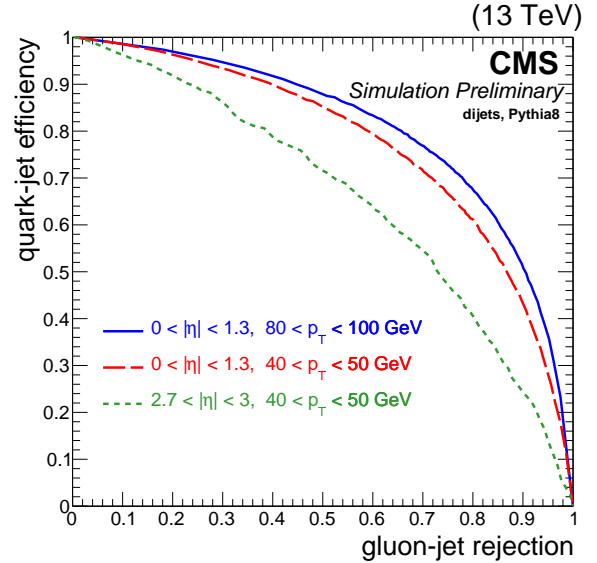
Figure 18: The CMS q/g tagging performance in simulation for two bins in jet p_T and two bins in jet $|\eta|$. Reproduced from Ref. [59].
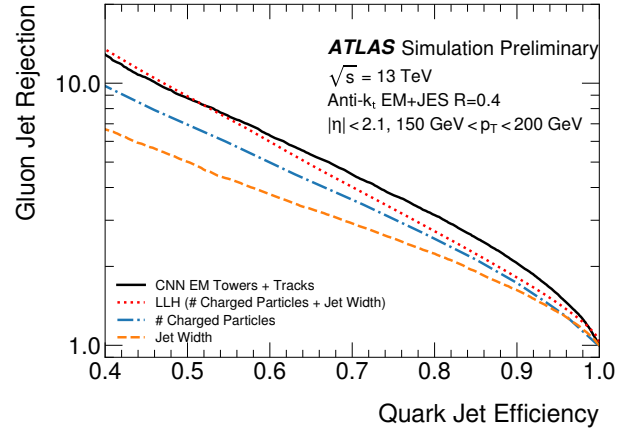
Figure 19: ROC curves for q/g taggers in ATLAS using physics motivated observables and modern machine learning techniques. Adapted from [99].

the dashed and dotted lines). Pythia tends to overestimate the q/g tagging performance with respect to data, as illustrated quite strikingly in Figure 21. In contrast, Herwig tends to underestimate the performance (leading to the famous 'Pythia-data-Herwig sandwich').

Data-driven reweighting techniques can be used to mitigate the observed mis-modeling (solid and open symbols

14
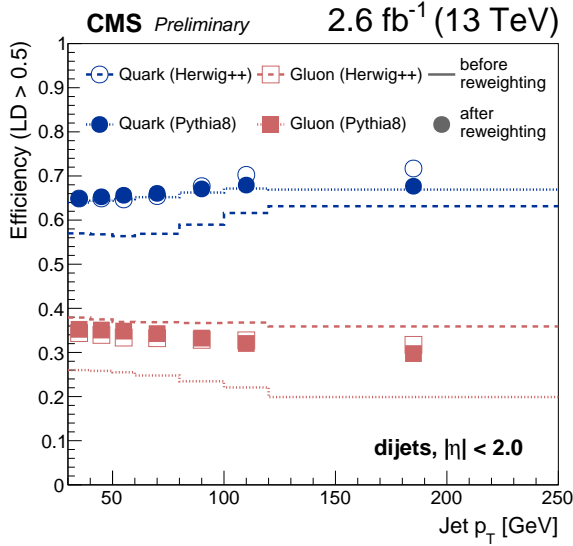
Figure 20: The CMS q/g tagging efficiency for a fixed working point of the tagger. The dashed (dotted) lines show the Herwig (Pythia) performance before reweighting, while the symbols show the performance after reweighting for quarks in blue and gluons in red. Reproduced from Ref. [59].

Figure 21: The two-dimensional q/g likelihood with AT-LAS data (left) and simulation (right). Reproduced from Ref. [96].

in Fig. 20). In particular, multiple samples with a different (but known) q/g composition can be used to extract the distribution of q/g tagging observables. ATLAS and CMS have both used Z/γ+jets and dijet samples, which are enriched in quark and gluon jets, respectively.
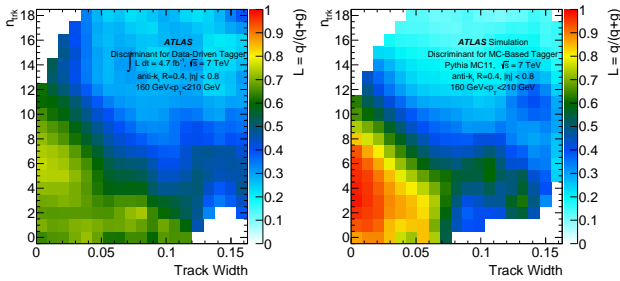
The Run 2 ATLAS tagger is based entirely on dijets, exploiting the rapidity dependence of the q/g fraction to extract the track multiplicity separately for quarks and
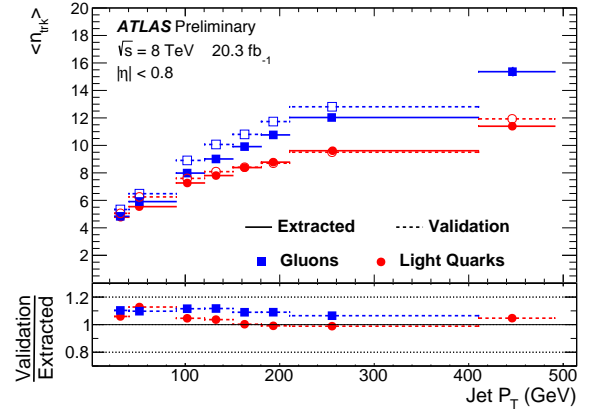

Figure 22: The average track multiplicity in ATLAS for Z/γ+jets (quark-enriched) and dijets (gluon-enriched). The dashed lines indicate the measurement on the validation samples : Z/γ+2-jets (quark-enriched) and trijets (gluon-enriched). Reproduced from Ref. [97].

gluons. A Run 1 measurement is used to constrain the particle-level modeling, and dedicated track reconstruction uncertainties are used to complement the particle-level uncertainty with a Run 2 detector-level uncertainty. The uncertainties on q/g tagging are 2-5% over a wide range of $200\,\mathrm{GeV} \lesssim p_T \lesssim 1\,\mathrm{TeV}$ at a working point of 60% quark jet efficiency [98]. The template-based calibration can also be used to directly construct the q/g tagger in data, avoiding mis-modeling concerns; however, when more than two observables are used to construct the tagger, it becomes impossible in practice to extract the high-dimensional templates.

Despite its power, the template technique has some residual non-closure because the resulting calibrated tagger applied to another final state may not have the same performance. This is illustrated in Fig. 22, which shows how the average track multiplicities extracted for quark and gluon jets (using high-purity Z/γ+jets and dijets data respectively) differ from the values obtained in the γ+2-jet and trijet samples used for validation.

Explicit tagging is often the focus of modern q/g discrimination, but there is a broad program of implicit tagging as well. One ubiquitous example of this is the ATLAS jet calibration procedure. Since the calorimeter response is non-linear, a jet with a higher particle multiplicity will have a lower response for the same energy. After applying a simulation-based correction to eliminate this inclusive bias in the JES (see Figure 6), a residual calibration is ap-

plied to correct for the dependence of the bias on the number of tracks associated to the jet and the jet width [33]. After applying this residual Global Sequential Calibration, the difference in response between quark and gluon jets is reduced. Implicit q/g tagging also appears in pileup jet identification [74, 75], boson and top tagging [27, 106, 108], and elsewhere.

Despite its long history, quark versus gluon jet tagging is still a very active topic of research. Since most analyses at the LHC target processes with a known and asymmetric q/g jet composition, q/g tagging holds great promise for improving searches and measurements in the future. Further studies are required to understand the limits of q/g tagging performance and to mitigate the sample dependence for universal definitions and calibrations. Interestingly, recent studies have shown how modern machine learning classifiers can be directly trained on data even though there are no per-jet labels [109, 110].

3.2 Vector Boson Tagging

The hadronic, two-prong decays of weak vector bosons V have a distinct radiation pattern compared to individual high-p_T quark or gluons; however, distinguishing V-jets from quark- or gluon-initiated jets requires a high-efficiency tagger across a range of p_T from 0.2 TeV up to multiple TeV, and $|\eta|$ from 0 to 2.5, while keeping the misidentification rate for quark and gluon jet backgrounds small. Stability of the misidentification rate with e.g. p_T and η allows data-driven background estimates to be used. This is necessary in light of the known mis-modeling in multijet Monte Carlo simulations. Good separation power between W and Z bosons is also desirable in a number of analyses, most notably searches for diboson resonances (see Sec. 5.2).

ATLAS and CMS performed a broad range of studies during Run 1 and the beginning of Run 2, systematically identifying the influence of pileup reduction and grooming techniques on jet substructure observables used for V tagging [37, 50]. Simulated samples containing W jets (rather than Z jets) are primarily used for these studies, as W jets are abundant in data thanks to the large quantity of $t\bar{t}$ events.

The optimization of the V-tagging algorithm is generally based on various factors concerning the tagged jet mass: (i) a sensible JMS (i.e., tagged jet mass close to the W mass), (ii) a narrow jet mass response with an approximate Gaussian lineshape, (iii) stability with respect to

pileup and jet p_T, and (iv) good background rejection at a given signal efficiency. Considering all of these factors, ATLAS decided on using the Trimming algorithm with $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.2$ on anti-k_t, $R = 1.0$ jets in Run 2, while CMS opted for using anti-k_t, $R = 0.8$ jets, treating the pileup first with PUPPI and then applying soft drop grooming with $z_{\text{cut}} = 0.1$ and $\beta = 0$.

In addition to the comprehensive studies of grooming options [27, 37, 38, 39, 111], ATLAS and CMS both investigated the discrimination powers for a plethora of jet substructure variables, including: N-subjettiness [60, 61], Qjet volatility [112], ratios of energy correlation functions C_2^β [91], D_2^β [113, 114] and N_2^β [115], angularities and planar flow [116], splitting scales [117, 118] the jet and subjet quark/gluon likelihood, and the jet pull angle [119].

Both ATLAS and CMS developed simple taggers that rely on the combination of the jet mass with one other variable that improves the discriminating power between the signal and background. The standard ATLAS V-tagger for Run 2 was chosen to be the trimmed jet ($R = 1.0, f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.2$) mass and $D_2^{\beta=1}$ [37], known as 'R2D2', while CMS opted to use the soft drop jet ($z_{\text{cut}} =0.1$, $\beta = 0$) mass and the N-subjettiness ratio $\tau_{21} = \tau_2/\tau_1$. Despite the different choices of tagging observables and detector design, ATLAS and CMS reach a very similar background rejection at a given tagging efficiency. An active field of developments is the usage of multivariate techniques for boosted V identification which have shown to be able to significantly improve the background rejection [50, 120].

In the ATLAS studies the variable $C_2^{\beta=1}$ in combination with the trimmed jet mass has been shown to be as good a discriminator as τ_{21}[3] as shown in Fig. 23. This is in contradiction to the study by CMS, where C_2^β is one of the weaker observables; however, a direct comparison is difficult, since in ATLAS *groomed* substructure variables are used, calculated for trimmed jets, while in CMS ungroomed variables are used. Also, the particulars of particle reconstruction have a large impact on the performance of individual observables. While a study of the performance of D_2^β at CMS is still pending, the soft drop N_2^β observable was found to give similar performance to τ_{21} in CMS [122].

CMS studied the quark/gluon likelihood (QGL) discriminator for its potential in V-tagging applications in

---

[3] A different axis definition for the subjet axes is used in ATLAS when calculating τ_N, known as *the-winner-takes-all* axis [121], which is consistently found to perform slightly better than the standard subjet axis definition in tagging bosons.
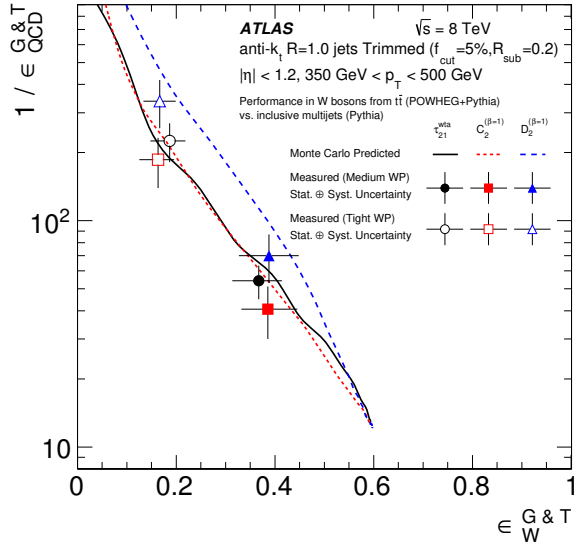
Figure 23: Signal efficiency versus background rejection power compared with measurements from ATLAS for $350 < p_T < 500\,\text{GeV}$. Taken from Ref. [37].

Run 1 [27], finding that a combination of the groomed jet mass and the QGL achieved a similar discrimination power as the groomed jet mass and τ_{21}. When adding the QGL to the Run 1 V-tagger (pruned jet mass and τ_{21}), the misidentification rate was reduced slightly from 2.6% to 2.3% at a constant signal efficiency of 50%. A similar reduction of the misidentification rate was observed when adding $C_2^{\beta=2}$, showing that C_2^{β} carries additional information with respect to the groomed jet mass and τ_{21}. However, the QGL and C_2^{β} exhibit a considerable pileup dependence, resulting in a degradation of their discrimination power with increasing activity. This pileup dependence is expected to be reduced when using PUPPI in place of particle flow + CHS.

In Fig. 23 the ATLAS measurements of signal efficiencies versus background rejection power are shown for τ_{21}, C_2^{β} and D_2^{β}, together with a selection on the trimmed jet mass (in this p_T range, the smallest mass window that captured 68% of the signal jets was found to be 71-91 GeV- see Ref. [37], table 7). The measurements are shown with statistical and systematic uncertainties. It is reassuring that the points for all three observables lie on the predicted performance curves for the two different working points studied.

In the ATLAS measurement the dominating systematic
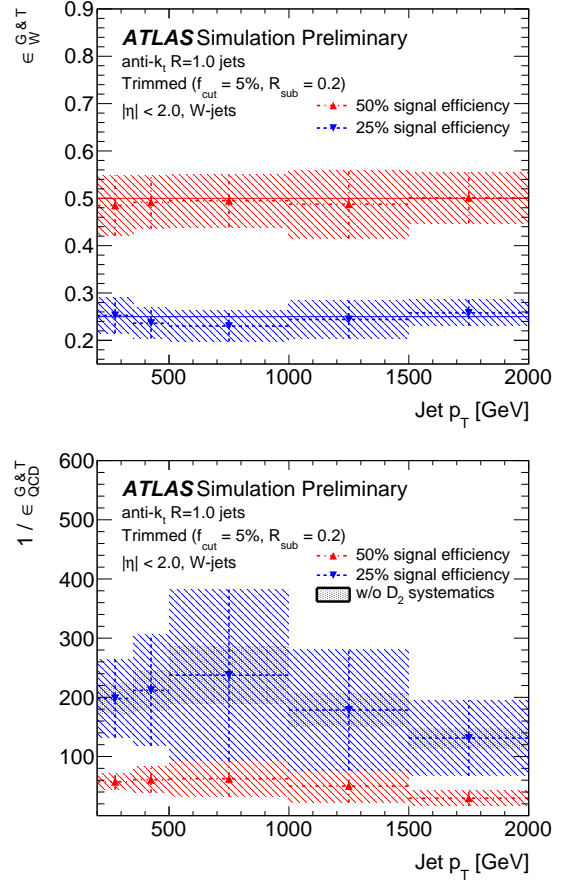
Figure 24: Efficiency (top) and misidentification rate (bottom) for tagging boosted W bosons in ATLAS. Adapted from [38].

uncertainty comes from the jet substructure scale, which has been derived by comparing calorimeter-jets with track-jets. Once again, the distributions in data lie between the ones derived with Pythia and Herwig, leading to large modeling uncertainties [37, 38]. A similar observation is made by CMS [50, 59]. Improving the modeling of jet properties and thereby reducing the differences between different event generators is a major task, but crucial for future precision studies using jet substructure.

A crucial aspect of V-tagging is the derivation of background rates from multijet production in real collision data when performing measurements. A commonly used method is the extrapolation from one or more control regions, which are defined orthogonally to the signal region. Usually, these control regions are defined by inverting the jet mass window selection, see e.g. [123, 124, 125, 126,

17

[127, 128]. Transfer functions are derived from simulation, extrapolating the rates and shapes from the control to the signal regions. Even though these transfer functions are ratios of distributions, which results in a reduction of the impact of modeling uncertainties, a residual dependence on the simulation can not be eliminated. However, the uncertainties in the high-p_T tails of the transfer functions can be eliminated by ensuring a constant behavior as a function of p_T. The requirement is thus a flat signal or background efficiency (depending on the needs of the analysis). In order to achieve a flat signal efficiency, ATLAS developed a p_T-dependent selection on the value of $D_2^{\beta=1}$, as this distribution shows a strong dependence on p_T [38]. In contrast to the Run 1 studies described above, no p_T-dependent selection is made on the trimmed jet mass, as the calibrated jet mass is used to define the V-tagging working point. While the jet mass resolution still increases with p_T, a constant window of $\pm 15\,\mathrm{GeV}$ around the mean reconstructed W or Z boson mass is used. This results in a p_T-dependent signal and background efficiency, which can also be countered with the p_T-dependent cut on $D_2^{\beta=1}$. This leads to a constant signal efficiency, while the background efficiency shows a residual p_T dependence, as shown in Fig. 24.

Another possibility has been explored by CMS. Instead of introducing p_T-dependent selection criteria, a linear transformation of the ratio τ_{21} has been studied [59], given by $\tau_{21}^{\mathrm{DDT}} = \tau_{21} - M \cdot \log(m^2/p_\mathrm{T}/1\,\mathrm{GeV})$ [129], where M is a constant determined from simulation. The replacement of τ_{21} with τ_{21}^{DDT} does not affect the overall performance of the tagger, but results in an approximately flat misidentification rate as a function of p_T, as shown in Fig. 25 (bottom), while introducing an increasing p_T-dependence for the V-tagging efficiency, as shown in Fig. 25 (bottom). The development of decorrelated jet substructure taggers is an active field with new techniques e.g. described in Refs. [130, 131, 132].

A less-studied possibility to lift the p_T-dependence of substructure observables is the application of variable-R jets [133]. By shifting the p_T-dependence to the jet-clustering level with a distance parameter proportional to p_T^{-1}, a stable position of the jet mass and jet substructure variables with respect to changes in p_T can be achieved [134]. This can lead to a stable tagging performance without the necessity of p_T-dependent optimization steps, but further experimental studies are needed to commission this strategy for use in analyses.

For some analyses the requirement of $p_\mathrm{T} \gtrsim 200\,\mathrm{GeV}$ is
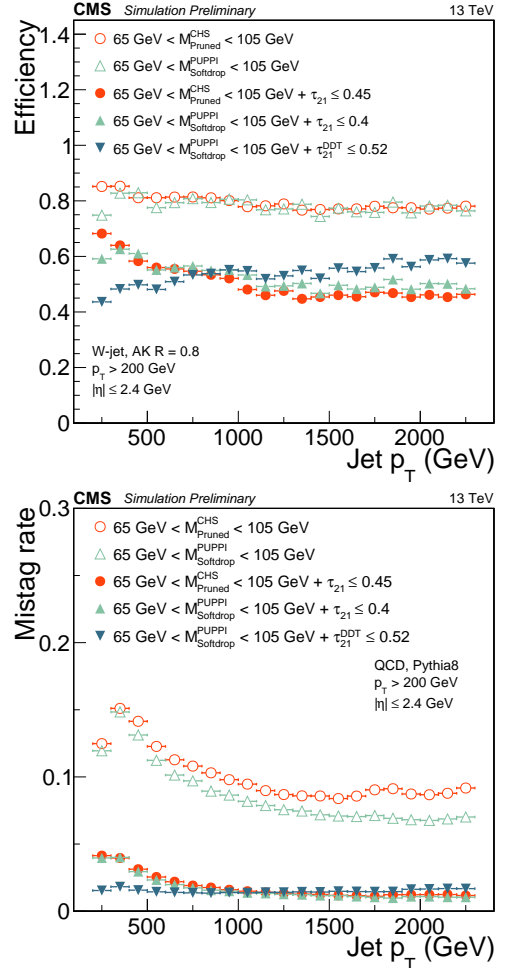
Figure 25: Efficiency and misidentification rate of various identification techniques for boosted W tagging. Taken from Ref. [59].

too restrictive, and hadronically decaying V bosons with lower p_T need to be selected. This poses a particular challenge due to the abundance of light flavor jets at the LHC and their indistinguishability from jets from W/Z decays. An attempt was made by CMS to discriminate 'resolved' (non-merged) hadronic W decays from multijet background using the QGL, the sum of the jet charges of the dijet pair and the jet pull angle. Combining these variables into a Boosted Decision Tree, a misidentification rate of about 25% is achieved for a signal efficiency of 50% [27]. While this is a first success, the performance is about an order of magnitude worse than V-tagging for fully merged decays, showing the power of substructure techniques in

Figure 26: Background rejection versus efficiency for discriminating Z boson jets from W boson jets for various jet observables and their combinations. Reproduced from Ref. [135].

this field.

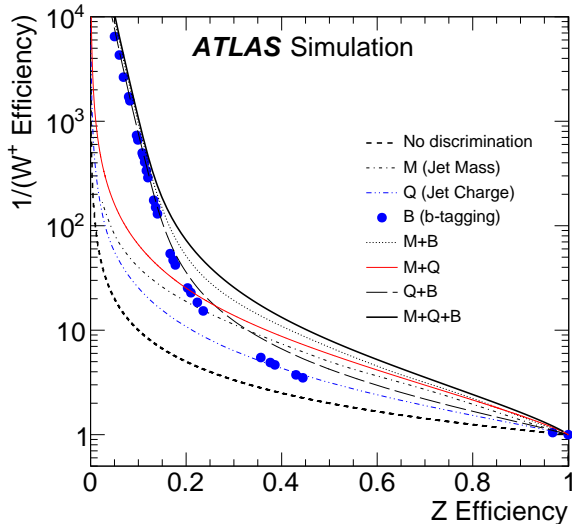In addition to developing tools for distinguishing boosted hadronically decaying W and Z bosons from generic quark and gluon jets, ATLAS has also built a tagger to further classify a boson jet as either originating from a W boson or a Z boson [135]. While theoretically clean due to the color singlet nature of the W and Z boson, this task is particularly challenging because the jet mass resolution is comparable to the difference $m_Z - m_W$. In order to improve the sensitivity of the tagger, jet charge and b-tagging information are combined with the jet mass. The jet mass distribution depends on the type of W or Z decay due to semi-leptonic B and D decays, so a full likelihood tagger is constructed by summing over the conditional likelihoods for each flavor type. To maximize the discrimination power from b-tagging, multiple efficiency working points are used simultaneously in the tagger. Figure 26 illustrates the performance of the boson type-tagger in simulation. A W^+ rejection near 8 (corresponding to a misidentification rate of 12.5%) is achieved at a Z boson efficiency of 50%. At this moderate Z boson efficiency, all of the inputs offer useful discrimination information. At low efficiencies, below the $b\bar{b}$ branching ratio for Z bosons, b-tagging dominates over the jet mass and jet charge.

The boson type-tagger was optimized for a relatively low boson boost, $200 < p_T < 400\,\text{GeV}$. The discrimination power of all of the input variables degrades with p_T due to the worsening jet mass resolution, tracking efficiency and momentum resolution, as well b-tagging efficiency. However, there are recent developments to address each of these challenges, such as the track-assisted jet mass (Sec. 1.5), pixel-cluster splitting [19], and track-jet b-tagging [136].

3.3 Top Tagging

The three-prong decays of highly boosted top quarks in the fully hadronic decay channel offer richer phenomenology for their identification than the two-prong decays of $W/Z/H$ bosons. This has been exploited in a number of algorithms, which usually aim at an optimal performance in a particular kinematic regime.

The techniques for tagging boosted top quarks have evolved as fairly complex methods in comparison to the V-taggers; these techniques include:

(a) The Johns Hopkins / CMS top tagger (CMSTT) [137] was designed for tagging top quarks with $p_T > 1\,\text{TeV}$. The algorithm is based on a decomposition of the primary jet into up to four subjets by reversing the Cambridge/Aachen clustering sequence. It has been adapted by the CMS Collaboration [138, 139], and was adopted as the standard top-tagging algorithm in CMS in Run 1, where it was typically used in the region of $p_T > 400\,\text{GeV}$, with an average identification efficiency of 38% at 3% misidentification rate [51].

(b) The HEPTopTagger (HTT) [140, 141] was designed to target $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel. In $t\bar{t}H$ production the top quark p_T distribution peaks around $150\,\text{GeV}$ and is steeply falling towards increasing p_T, where it is already an order of magnitude smaller at $p_T \sim 400\,\text{GeV}$. This results in a requirement of non-zero signal efficiency already at $p_T \approx 200\,\text{GeV}$, where the top quark decay is only moderately boosted. The HTT achieves this with a large jet distance parameter of 1.5 and a sequence of declustering, filtering and re-clustering of the original Cambridge/Aachen jet. The performance of the HTT was studied by the ATLAS and CMS Collaborations on data with a center-of-mass energy $\sqrt{s} = 7$ and 8 TeV [40, 51, 142]. Efficiencies of 10% with misidentification rates of 0.5% for jets with

$200 < p_T < 250$ GeV were observed. The efficiency increases with increasing jet p_T, where a plateau is reached for $p_T > 400$ GeV, with efficiencies of approximately 40% at 3% misidentification rate, very similar to the performance achieved with the CMSTT.

(c) Shower Deconstruction [143, 144] was designed to be analogous to running a parton shower Monte Carlo generator in reverse, where emission and decay probabilities at each vertex, color connections, and kinematic requirements are considered. Small-radius (generally $R = 0.2$) subjets are reconstructed with the Cambridge/Aachen algorithm and all possible *shower histories* that can lead to the observed leading final state anti-k_t, $R = 1.0$ jet are calculated. Each shower history is assigned a probability weight factor based on the aforementioned considerations (to be signal-like or background-like), then a likelihood ratio $\chi(p_N)$ is constructed, and the $\log \chi(p_N)$ is used as the discriminating substructure variable. For top quark tagging, efficiencies of 80% with misidentification rates of 50% for jets with $500 < p_T < 1000$ GeV were observed. The efficiency increases with increasing jet p_T, where a plateau is reached for $p_T > 2000$ GeV, with efficiencies of $\sim 80\%$ at 10% misidentification rate. Recently, the Shower Deconstruction algorithm was optimized for top quarks with $p_T > 800$ GeV in context of the W' to tb hadronic search [145] by using exclusive k_T subjets.

In addition to the dedicated techniques described above, simpler algorithms using grooming and substructure as per V-tagging have been investigated by ATLAS for a number of years. A performance study at 7 TeV [142] investigated a variety of performance metrics relating to the usage of groomed jets. Different grooming algorithms were investigated for their resilience to pileup and mass resolution. It was concluded that trimmed anti-k_t jets with a distance parameter of 1.0 and trimming parameters of $R_{sub} = 0.3$ and $f_{cut} = 0.05$ were a very good candidate for a one-fits-all large-R jet definition. This jet definition became standard in ATLAS for $W/Z/H$ and top quark tagging in Run 1. A later ATLAS study [40] investigated the various methods available for tagging hadronic, highly boosted top quarks. The so-called Tagger V has $M_{jet} > 100$ GeV, $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 20$ GeV, where $\sqrt{d_{ij}}$ is the k_T-splitting scale [117]. The efficiency versus rejection is shown for various taggers in Fig. 27. The difference between Taggers III and V is the additional requirement on
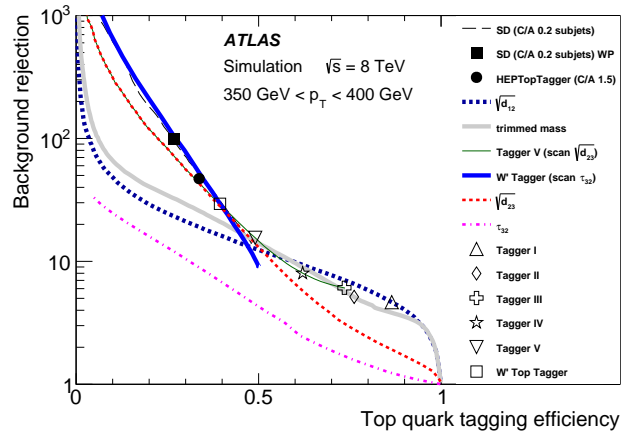

Figure 27: Top quark tagging efficiency versus background rejection for various substructure variables and combinations in ATLAS. Taken from Ref. [40].

$\sqrt{d_{23}}$ in Tagger V. At efficiencies smaller than 45%, the W' tagger, based on $\sqrt{d_{12}}$ and the N-subjettiness ratios τ_{21} and τ_{32}, has better background rejection than Taggers III and V. ATLAS also tested the HTT and Shower Deconstruction [146], which have been found to have good background rejection (larger than 50) for efficiency values smaller than about 35%. However, similar as for the CMS experiment, the background efficiencies of the two taggers show a significant rise with increasing p_T.

CMS has focused on enhancing the performance of CMSTT and HTT by identifying observables which carry discriminatory power, but have only small or moderate correlations with the observables used in the main algorithm. Typically, correlation coefficients of about 0.3 or less are required for noticeable improvement when augmenting an algorithm with additional variables. Examples for discriminating variables which fulfill this are N-subjettiness ratios, energy correlation functions and their ratios, and b-tagging. A study by the CMS Collaboration showed that at 20% signal efficiency, the background rejection of the CMSTT could be improved by a factor of 5 when adding information from τ_{32} and subjet b-tagging [51]. At higher efficiencies, the improvements become smaller. For the HTT, improvements of similar size were observed for $p_T > 200$ GeV, becoming less significant at higher p_T.

The ATLAS choice of $R = 1.0$ jets compared to CMS ($R = 0.8$) results in an earlier rise of the tagging efficiency with increasing jet p_T.

The large difference in performance of the single variable

τ_{32} between ATLAS (Fig. 27) and CMS (Fig. 28) is due to jet grooming. Although the CMS study shows only the ROC curves for $800 < p_T < 1000\,\mathrm{GeV}$, the overall picture does not change when studying top quarks in the region of $p_T \approx 400\,\mathrm{GeV}$. Instead, in ATLAS τ_{32} is calculated from trimmed jets, which results in less discrimination power when used as sole tagging variable compared to ungroomed τ_{32}. However, groomed τ_{32} can still lead to considerable improvements when combined with other variables.

As with V-tagging discussed above, ATLAS and CMS took advantage of the LHC shutdown between Run 1 and Run 2 to perform broad studies of the different top-taggers available, with emphasis on their stability with respect to pileup and other detector effects, instead of the utmost gain in performance [41, 52]. Single variables and their combinations were studied and compared with Shower Deconstruction, CMSTT, HTT, and an improved version of the HTT with shrinking cone size (HTTv2) [147].

Figure 28 shows a comparison based on simulation of the single variable performance in CMS, where signal jets were generated through a heavy resonance decaying to $t\bar{t}$ and background jets are taken from QCD multijet production. Note that for this study reconstructed jets are matched to a generated parton, and the distance between the top quark and its decay products must be less than 0.6 (0.8) for a reconstructed $R = 0.8$ (1.5) jet, to ensure that the top quark decay products are fully merged and reconstructed in a single jet. The best single variable in terms of efficiency versus background rejection is the discriminator $\log \chi$, calculated with Shower Deconstruction. The second best variables are the N-subjettiness ratio τ_{32} at low efficiency and the jet mass calculated with the HTTv2 at high efficiency values. The individual groomed jet masses show similar performance, and the CMS Collaboration moved to using the soft drop mass due to its beneficial theoretical properties [7]. The default for CMS Run 2 analyses was chosen to be the soft drop jet mass combined with τ_3/τ_2 for top tagging at high p_T. Generally, at high boost, the combination of a groomed mass with τ_{32} leads to a large gain in background rejection.

The CMS study also investigated combining single variables with more complex taggers. Combining Shower Deconstruction with the soft drop mass, τ_{32}, and subjet b-tagging can lead to improvements, as shown in Fig. 29; however, the efficiency and misidentification rate for this combination were found not to be stable as a function of jet p_T (the combined algorithms were studied using working points corresponding to a background efficiency of 0.3).
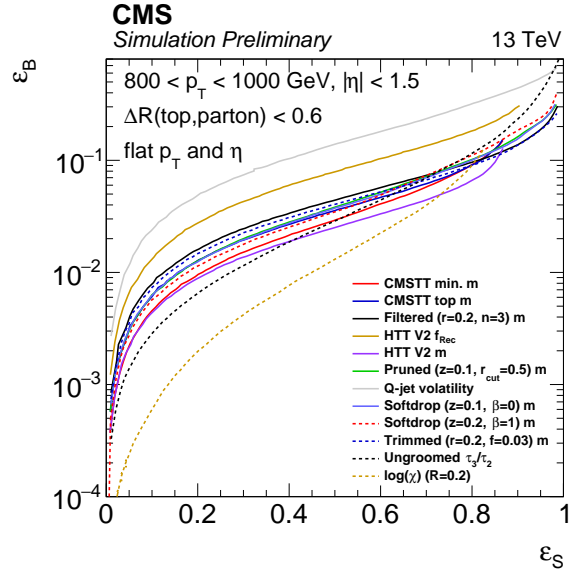

Figure 28: Background versus signal efficiency for the single variables studied in the optimization of top tagging for CMS Run 2 analyses. Taken from Ref. [52]

At low boosts, the dedicated HTTv2 shows the best performance. In this kinematic region, using groomed τ_{32}, obtained by using the set of particles from the soft drop jet instead of the original jet, helps to improve the performance.

In the shutdown between Run 1 and Run 2, ATLAS commissioned a single top tagger for use by physics analyses. The rationale behind this approach was the potential benefit of having an efficient top tagger with well-understood efficiency and associated systematic uncertainties validated in the Run 1 dataset. Similarly as for Run 1, the supported top-tagger makes use of anti-k_t, $R = 1.0$ trimmed jets, but with a parameter of $R_{sub} = 0.2$ instead of 0.3 as used in Run 1. Candidate top jets are required to satisfy a calibrated mass window requirement $122.5 < M_{jet} < 222.5\,\mathrm{GeV}$ and a p_T-dependent, one-sided cut on τ_{32} [41]. The background rejection for fixed signal efficiency working points for the pairwise variable combinations considered is shown in Ref. 30. The ratio τ_{32} was used instead of the previously used k_T-splitting scales or the minimum pairwise mass Q_W [118] because of a smaller correlation with M_{jet} (shown in Ref. 31), a reduced p_T-dependence, and good performance of this combination across a large range in p_T. The distribution of τ_{32} mea-
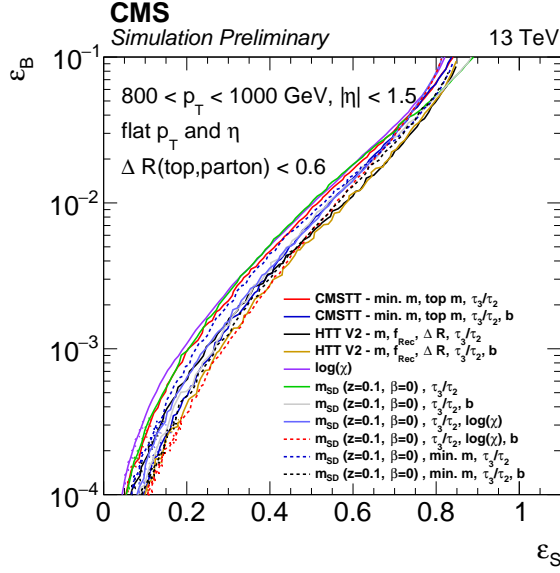
Figure 29: Background efficiency versus signal efficiency for combined variables studied in the optimization of top tagging for CMS Run 2 analyses. Taken from Ref. [52]

Figure 30: The ATLAS pre-Run 2 study showing the background rejection obtained for two signal efficiency points (50% and 80%), for jets with $200 < p_T < 500\,\mathrm{GeV}$. The statistical uncertainties are around 1 unit of background rejection. Adapted from [41].

sured in dijet events with Run 2 data is shown in Fig. 32.

A common problem of top-tagging algorithms is the rise of the misidentification rate with increasing p_T, which is due to the peak of the mass distribution for quark- and gluon-initiated background jets shifting to higher values. For some taggers, for example the CMSTT, this shift also results in a decrease of the efficiency once a very high p_T threshold is crossed (larger than $1\,\mathrm{TeV}$) [139]. A possible solution to this is offered by the variable-R (VR) algorithm, introduced in Sec. 3.2. The ATLAS Collaboration studied the performance of the VR algorithm for top-tagging and reported a stabilization of the position of the jet mass peak for a large range of p_T [134]. The VR jets are shown to improve the performance of the jet mass, $\sqrt{d_{12}}$ and τ_{32} for top tagging, when compared to trimmed jets. An interesting development using VR jets is the Heavy Object Tagger with variable-R (HOTVR) [149], which combines the VR algorithm with a clustering veto, resulting in a single jet clustering sequence producing groomed jets with subjets.

Most top-taggers target either the region of low to intermediate boosts, or the highly boosted regime. However, in typical searches for new physics at the LHC non-vanishing efficiency for the full kinematic reach is crucial. Several
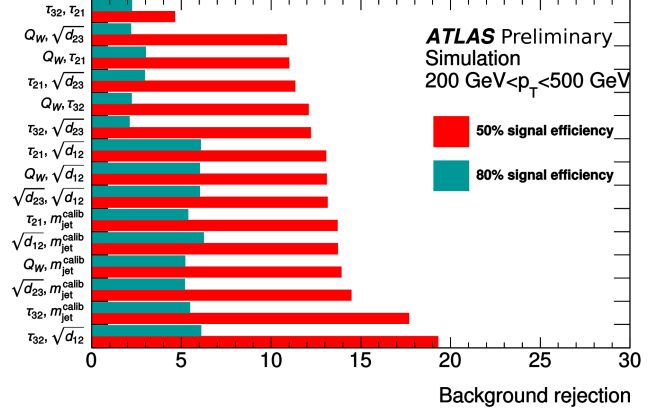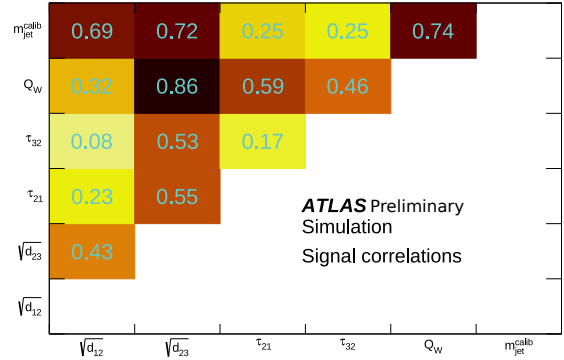

Figure 31: The ATLAS pre-Run 2 study showing the linear correlation coefficients between variables, for jets with $200 < p_T < 500\,\mathrm{GeV}$. Adapted from Ref. [41].
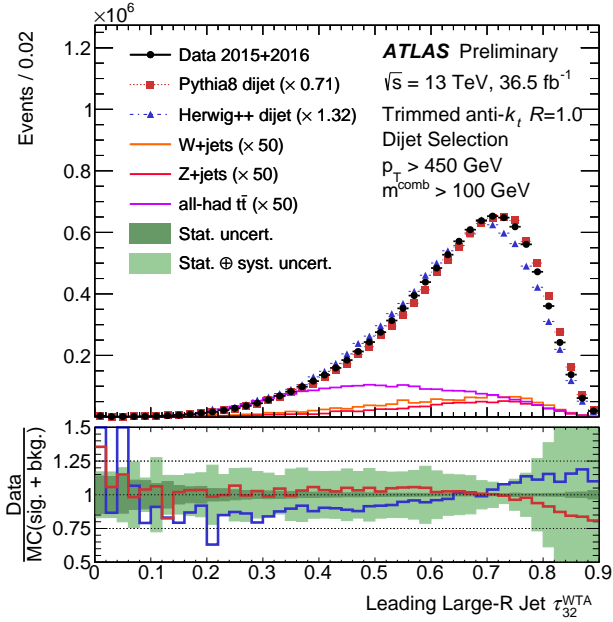
Figure 32: Measured distribution of the N-subjettiness ratio τ_{32} calculated on trimmed jets for a dijet selection with $p_T > 450\,\text{GeV}$ and $p_T > 200\,\text{GeV}$ for the leading and subleading jet, respectively. The data are compared to simulated events, where the dijet samples have been normalized to the signal-subtracted data. Taken from Ref. [148].

attempts of combining different reconstruction and identification algorithms have been made. A search for resonances decaying to $t\bar{t}$ by the ATLAS Collaboration uses a cascading selection from boosted to resolved [150], where the resolved topology is reconstructed and identified using a χ^2-sorting algorithm. To efficiently identify top quarks over a broad p_T range in the search for top squark pair production, reclustered variable-R jets are used with $R = 0.4$ jets as inputs to the jet reclustering algorithm [47, 151].

A search for supersymmetry in CMS [152] uses three distinct topologies: fully-merged top quark decays (Monojet), merged W boson decays (Dijet) and resolved decays (Trijet). The efficiency of the three categories is shown in Fig. 33, where the turn-on of the combined efficiency starts at values as low as $p_T \approx 100\,\text{GeV}$. The resolved trijet category is identified using three anti-k_t jets with a distance parameter of 0.4, where the large combinatorial background is suppressed through a multivariate analysis, which achieves a misidentification rate of approximately 20%. There exist other approaches to cover the transition from low to high Lorentz boosts, using a single algorithm.
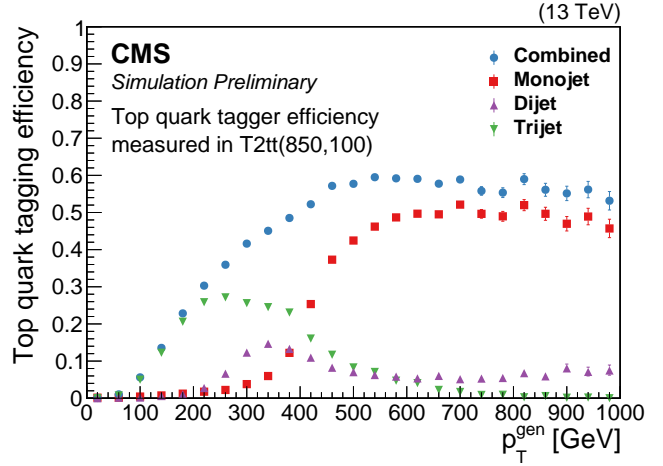

Figure 33: Top tagging efficiency of three different top tagging methods and the combined efficiency, as a function of the generated top quark p_T. Taken from Ref. [152].

In the HTTv2 algorithm, the jet size is reduced until an optimal size R_{opt} is found, defined by the fractional jet mass contained in the smaller jet. This results in better performance at high p_T, while keeping a low misidentification rate at low p_T. The distribution in the variable R_{opt} is shown in Fig. 34 for jets with $p_T > 200\,\text{GeV}$, where R_{opt} is observed to vary between 0.4 and 1.5.

An important step towards the commissioning of top taggers within an experiment are measurements of the efficiency and misidentification rate in real collision data. Generally, high-purity samples of top-jets in data are obtained using a tight signal selection (an electron or muon, well-separated from a high-p_T large-R jet, and an additional b-tagged jet) to ensure that events contain a fully-merged top quark decay in a single large-R jet. This can never be fully achieved, as no requirements on the substructure of the large-R jet can be imposed without biasing the efficiency measurement. This results in an efficiency measurement that will be based on a sample also containing partially-merged or even non-merged top quark decays. These can be subtracted from the efficiency measurement by using simulated events, as done in a study by the ATLAS Collaboration [40], with the drawback of relying on a specific simulation and the ambiguous definition of a fully-merged top quark decay. By not correcting for non-merged top quark decays, efficiency values are obtained smaller than the ones suggested by ROC curve studies, see for example [52]. The optimal way of obtain-
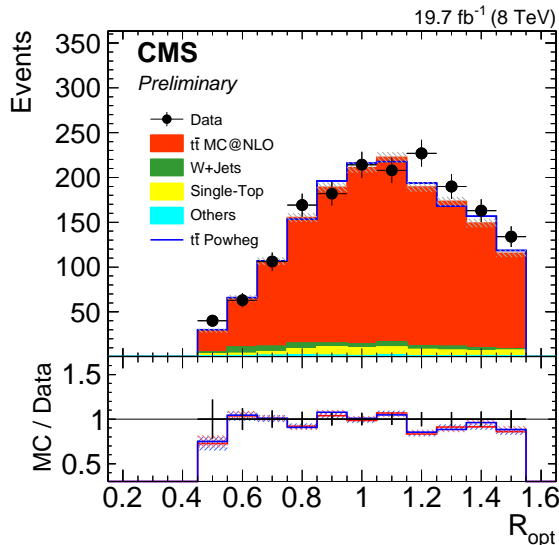
Figure 34: Distribution of the optimal jet size R_{opt} of the HTTv2, as measured by the CMS Collaboration. Taken from Ref. [52].

ing an efficiency measurement that relies only minimally on simulation and does not rely on the subtraction of processes with similar features to the signal is to fit the fully- and partially-merged categories simultaneously, as done in a recent measurement by CMS [153].

Measurements of the misidentification rate can be carried out by selecting a dijet sample, which is dominated by light-flavor jets. The disadvantage of this approach is the high p_{T} threshold of unprescaled jet triggers, which results in measurements starting from $p_{\text{T}} > 400\,\text{GeV}$ or higher. A solution to this is the tag-and-probe method, in which the tagged jet can be required to fail top-tagging selection criteria, resulting in a sample with negligible contamination of $t\bar{t}$ production, even after requiring the probe jet to be top-tagged [52]. Another approach is to use a non-isolated electron trigger, where the electron fails offline identification criteria. This yields events mainly from light-flavor multijet production, where a jet is misidentified as an electron at the trigger level. While the top-tag misidentification rate can be measured starting from smaller values of p_{T} with this strategy, a non-negligible amount of $t\bar{t}$ contamination has to be subtracted after requiring a top-tagged jet [40].

As an example, the efficiency and misidentification rate of Shower Deconstruction with the requirement $\log(\chi) > 2.5$, as measured in ATLAS, are shown in Fig. 35. The
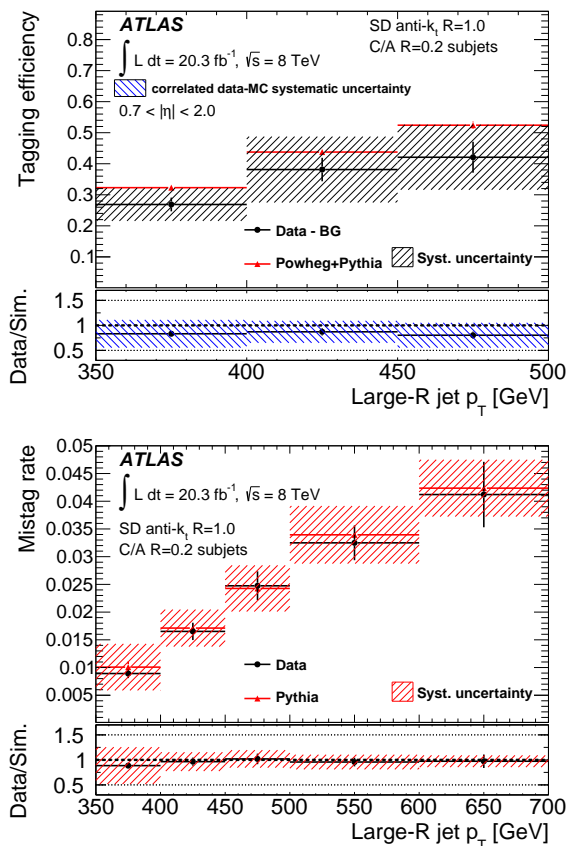
Figure 35: ATLAS measurement of the efficiency (top) and misidentification rate (bottom) for trimmed jets with a distance parameter of 1.0 tagged with Shower Deconstruction. Taken from Ref. [40].

efficiency of 30% with a misidentification rate of 1% for $350 < p_{\text{T}} < 400\,\text{GeV}$ agrees well with the values obtained from Fig. 27. Note that the largest uncertainty of the efficiency measurement stems from the choice of the MC generator used to simulate $t\bar{t}$ production. The uncertainty of the misidentification rate measurement is dominated by the energy scales and resolutions of the subjets and large-R jets.

3.4 $H \to b\bar{b}$ Tagging

The ability to identify jets originating from b quarks produced in the decay of boosted heavy particles, such as top quarks or Higgs bosons, is critical for measurements and searches of new physics signatures, especially as the LHC moves to higher energies.

ATLAS and CMS both use dedicated b-tagging algorithms that have been developed and optimized over more than a decade. Both experiments use multivariate techniques making use of the long lifetime of b hadrons. For Run 2 analyses, CMS uses the CSVv2 algorithm [154] and ATLAS uses the MV2c10 algorithm [155].

The boosted $H \to b\bar{b}$ signal represents a good signature for many different scenarios beyond the Standard Model: resonant HH and VH production, searches for boosted mono-H, or vector-like quark searches in the tH and bH final states. Because of the large predicted branching fraction for the $H \to b\bar{b}$ decay of about 58%, its coupling to b quarks is one of the most interesting to study. For a large fraction of Higgs bosons with $p_T > 300$ GeV, the two b quark jets merge into a single jet for a jet distance parameter of $R = 0.8$ or 1.0, as used in CMS and ATLAS respectively. The approach to reconstruct the Higgs boson in this topology is different from considering the individual smaller-sized jets separately, as the composite nature of such a jet is revealed by analyzing its substructure. Several phenomenological studies have explored $H \to b\bar{b}$ tagging algorithms using jet substructure, though ultimately the optimal performance comes from using a combination of substructure information and the track and vertex information related to the b hadron lifetime.

The approaches to identify boosted $H \to b\bar{b}$ candidates that have been explored (and used) at CMS and ATLAS include:

(a) Subjet b-tagging [136, 156, 157, 158, 159, 160], where 'standard' b-tagging is applied to each of the subjets (the standard for CMS is the CSVv2 algorithm [154], and for ATLAS is MV2c20 [155]). Tagging b-jets in dense environments is of particular importance here, and was studied by ATLAS in Ref. [161]. In CMS subjets with $R = 0.4$ are clustered with the k_T algorithm using the constituents of the large-R jet, while for ATLAS track jets with a radius of 0.2 are matched to the large-R jet using the ghost-association technique. At high p_T the subjets start to overlap causing the standard b-tagging techniques to break down due to double-counting of tracks and secondary vertices when computing the subjet b-tag discriminants.

(b) Double-b tagging [160, 162, 163]. In ATLAS, the term double-b tagging means that the two leading p_T track jets must pass the same b-tagging requirement. In CMS, the double-b tagger is a novel approach detailed in [162, 163], which uses the pruned anti-k_t, $R = 0.8$
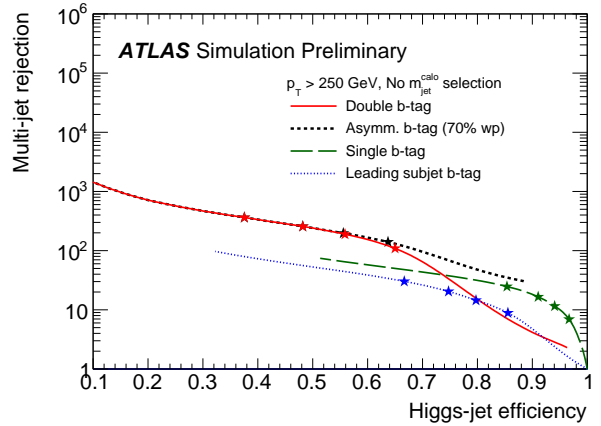

Figure 36: The rejection of inclusive multijets versus Higgs-jet efficiency using all large-R jets with $p_T > 250$ GeV for single, double, asymmetric, and leading subjet b-tagging requirements. Taken from Ref. [160].

jet mass (a window of $50 < M < 200$ GeV is required to reduce the multijet background) and the N-subjettiness axes.

The Higgs-jet efficiency versus the inclusive multijet rejection are shown in Fig. 36 for ATLAS subjet b-tagging, where the performance curves are shown for double-b tagging, leading subjet b-tagging, and asymmetric b-tagging (among the two leading p_T track jets, the track jet with the largest b-tagging weight must pass the fixed 70% b-tagging working point threshold, while the b-tagging requirement of the other jet is varied) requirements. None of the curves reach a Higgs-jet efficiency of 100% due to the imperfect efficiency to reconstruct the track jets needed for b-tagging and, in the case of asymmetric b-tagging, also due to the 70% b-tagging working point requirement on one of the track jets.

The CMS double-b tagging approach has been applied to identify $H \to b\bar{b}$ candidates [162, 163]. The algorithm attempts to fully exploit the strong correlations between the b hadron flight directions and the energy flows of the two subjets, while adapting the variables used in the CSVv2 algorithm. The flexibility of the double-b tagger is ensured by avoiding a strong performance dependency on the jet p_T and mass.

With the double-b tagger, at the same signal efficiency, the misidentification rate is uniformly lower by about a factor of two compared to the subjet b-tagging approach. Given the different kinematic properties expected for a

$b\bar{b}$ pair originating from the decay of a massive resonance compared to gluon splitting, the misidentification rate for the gluon splitting background reduces from 60% to 50% at 80% signal efficiency and from 20% to 10% at 35% signal efficiency. At high p_T, even larger performance improvements are observed, which is an important gain for searches for heavy resonances, where very high p_T jets are expected. In Fig. 37 the signal efficiencies and misidentification rates for the double-b tagger are shown as a function of jet p_T for three operating points: loose, medium and tight, which correspond to 80%, 70% and 35% signal efficiency, respectively, for a jet p_T of about 1000 GeV. The misidentification rate is mostly flat across the p_T range considered while the signal efficiency decreases with increasing p_T, as expected from the degradation of the tracking performance inside high p_T jets.
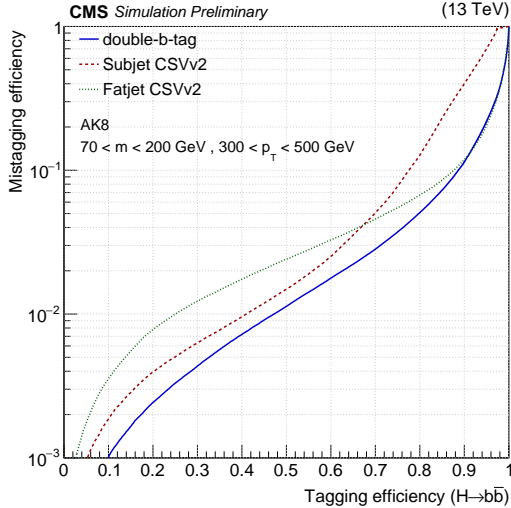

Figure 37: The misidentification efficiency for inclusive multijets versus Higgs-jet efficiency using jets with $300 < p_T < 500$ GeV and pruned jet mass $70 < m < 200$ GeV for three different b-tagging requirements. Taken from Ref. [162].

Due to the small cross section of producing events with boosted $H \to b\bar{b}$ or $Z \to b\bar{b}$ jets, the efficiency of the ATLAS and CMS Higgs identification algorithms is measured using QCD multijet events enriched in jets from gluon splitting, $g \to b\bar{b}$ with a topology similar to that of boosted $H \to b\bar{b}$ jets.

CMS selects topologies as similar as possible to a signal jet by requiring the jet $p_T > 300$ GeV and pruned

mass > 50 GeV [162, 163]. Each jet has to contain at least two muons, each with $p_T > 7$ GeV and $|\eta| < 2.4$. Each pruned subjet is required to have at least one muon among its constituents and within $\Delta R < 0.4$ from the subjet axis ("double-muon tagged"). An alternative selection that requires at least one muon is also examined as a cross-check for the measurement ("single-muon tagged"). While this single-muon selection allows for a larger dataset in which to perform the tagger efficiency measurement, the gluon splitting topology in this inclusive phase space is less signal-like relative to the double-muon selection. Thus, to maximize the similarity between the $g \to b\bar{b}$ and the $H \to b\bar{b}$ topologies, the measurement is performed requiring double-muon tagged jets. It is worth noting however that the jet mass depends on the number of muons and a large fraction of the signal will not contain two muons.

ATLAS performed a similar measurement selecting events with at least one anti-k_t, $R = 1.0$ jets with $p_T > 250$ GeV that has two ghost-associated $R = 0.2$ track jets [160]. As opposed to the measurement from CMS, only one of the subjets is required to have a muon associated to it. Kinematic and substructure variables are compared in data and MC after correcting for flavor composition differences of the large-R jet observed between data and MC simulation and are found to be in good agreement.

One of the major backgrounds for analyses selecting boosted H or Z bosons decaying to $b\bar{b}$ is $t\bar{t}$ production. The misidentification rate for boosted top quark jets faking H jets was measured in data by CMS [162, 163] in enriched data samples of lepton+jets $t\bar{t}$ events.

As previously discussed, for high p_T of the Higgs boson, the two subjets from b quarks start overlapping and the performance of identifying the subjets as fixed-radius track jets decreases significantly. To improve the performance of the ATLAS standard $H \to b\bar{b}$ identification algorithm for searches that are sensitive to high p_T Higgs bosons, the ATLAS Collaboration studied alternative methods: the use of variable-radius track jets, exclusive k_T subjets, calorimeter subjets reconstructed in the center-of-mass frame of the Higgs jet candidate [164] and the combination of three jet shape and jet substructure variables into a multivariate discriminator [165]. For highly boosted Higgs bosons, these reconstruction techniques significantly outperform the usage of fixed-radius track jets.

Table 2: Taggers in use by ATLAS and CMS for Run 2 of the LHC.

Q/G Taggers	Description
$LLH(n_{\text{track}}, W)$	Log Likelihood using track multiplicity and jet width
CNN	Convolutional Neural Network Tagger (Jet Images)
n_{track}	Simple n_{track} tagger
$LD(n_{cons}, \sigma_2, p_{\text{T}}D)$	Likelihood using constituent multiplicity, minor axis and fragmentation function
$BDT(n_{ch}, \sigma_1, \sigma_2, p_{\text{T}}D, W^x)$	BDT using charged multiplicity, minor and major axes, fragmentation function and 'ΔR-weighted p_{T} sum'

Vector Boson Taggers	Description
R2 D2	Trimmed ($R = 1.0$, $f_{\text{cut}} = 0.05$, $R_{\text{sub}} = 0.2$) mass and D_2^β
CMS V-tagger	PUPPI soft drop ($z_{\text{cut}} = 0.1$ and $\beta = 0$) mass and ungroomed τ_{21}
N_2^β	PUPPI soft drop ($z_{\text{cut}} = 0.1$ and $\beta = 0$) mass and soft drop N_2^β

Top Taggers	Description
R2 τ_{32}	Trimmed ($R = 1.0$, $f_{\text{cut}} = 0.05$, $R_{\text{sub}} = 0.2$) mass and τ_{32}
SD	Shower Deconstruction
HTT	HEPTopTagger
HTTv2	HEPTopTagger with shrinking cone
CMSTT	mMDT mass and τ_{32} and subjet b-tagging

Higgs Taggers	Description
CSVv2	Subjet b-tagging with the CSVv2 algorithm and $R_{\text{sub}} = 0.4$ jets
MV2c20	Subjet b-tagging with the MV2c20 algorithm and $R_{\text{sub}} = 0.2$ track jets
Double-b tagging	With the CSVv2 algorithm and pruned $R = 0.8$ jets, using N-subjettiness axes

4 Measurements

The measurement of jet properties is crucial to constrain the Standard Model in new energy regimes and constitutes and important test of perturbative calculations of jet structure over a wide region of phase space. Moreover jet cross section measurements provide constraints on the parton distribution functions and the strong coupling constant, α_s. The precise knowledge of jet properties also improves the precision of other measurements due to an improved modeling of important background processes and is necessary to probe for physics beyond the Standard Model. Jet substructure variable measurements are challenging as they require a precise measurement of the radiation pattern within the jet and thus the jet constituents.

4.1 Measurements with jet substructure

4.1.1 $t\bar{t}$ Cross Section

The selection cuts applied in the traditional $t\bar{t}$ cross section measurements [166, 167, 168, 169, 170, 171, 172] are chosen to maximize the acceptance and minimize the associated uncertainties on the fiducial cross section measurement. The resulting fiducial region is such that events with top p_T below $100\,\text{GeV}$ and above $600\,\text{GeV}$ are underrepresented, with the latter being due to collimated decays resulting from the large Lorentz boost. This is evident from Fig. 38, where a drop in selection efficiency below $100\,\text{GeV}$ and above $600\,\text{GeV}$ is apparent. This results in a small number of events being selected with high top quark p_T, as seen in the ATLAS Run 1 (7 TeV) measurement shown in Fig. 39; unfortunately, a very interesting region in terms of new physics is the least well-measured. Despite often having similar signal efficiencies to resolved reconstructed techniques, boosted top tagging techniques allow for more precise measurements at high p_T due to their higher background rejection.

The ATLAS Collaboration performed a measurement of the boosted $t\bar{t}$ differential cross section as a function of the p_T in the lepton+jets channel [173]. A least one anti-k_t jet, trimmed with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$ is required with $|\eta| < 2$ and $p_T > 300$ GeV. To select events with boosted top quarks, the large-R jet is required to have a mass larger than 100 GeV and $\sqrt{d_{12}} > 40$ GeV (Tagger III, see Sec. 3.3). The reconstructed p_T distribution of the anti-k_t $R = 1.0$ trimmed jet is unfolded to the parton and particle-level. The measured particle-level differential cross section is compared in Fig. 40 to the predictions of
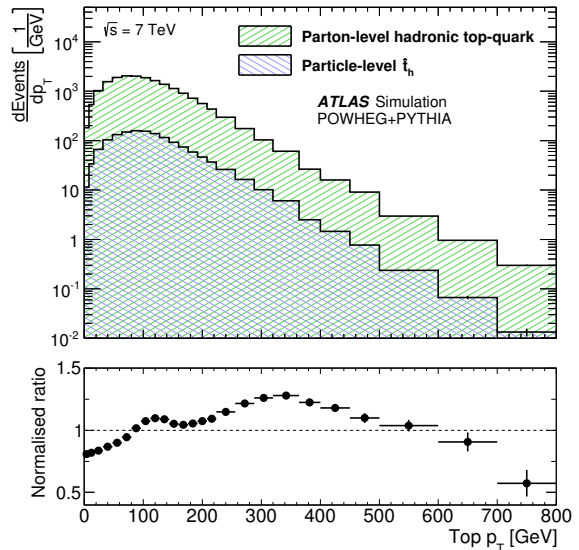

Figure 38: The drop in fiducial efficiency at top $p_T > 600$ GeV when reconstructing top quarks with individual anti-k_t, $R = 0.4$ jets (*resolved* reconstruction). Adapted from [169].

several MC generators normalized to the NNLO+NNLL inclusive cross section. Overall good agreement is observed, but a harder p_T spectrum is predicted by the simulation than observed in data with larger discrepancies at high p_T. The differential cross section measurement is also compared to predictions from Powheg+Pythia in Fig. 40 using either the HERAPDF or CT10 PDF set and two different values for h_{damp}, $h_{\text{damp}} = m_{\text{top}}$ and $h_{\text{damp}} = \infty$. The best data/MC agreement is observed when using the HERAPDF set and $h_{\text{damp}} = m_{\text{top}}$. For each of the settings, the trend of a harder p_T spectrum in simulation compared to data persists.

A similar measurement by the CMS Collaboration based on 8 TeV data [174] uses the CMSTT algorithm to reconstruct boosted top quarks. The unfolded results are in agreement with the ATLAS measurement and show a similar trend between data and simulation, as shown in Fig. 41.

These measurements extend up to a top quark p_T of 1.2 TeV, allowing for higher precision thanks to the usage of jet substructure techniques. The largest uncertainties at the highest values of p_T in ATLAS and CMS come from the large-R jet energy scale and the extrapolation of the b-jet calibration to high p_T.
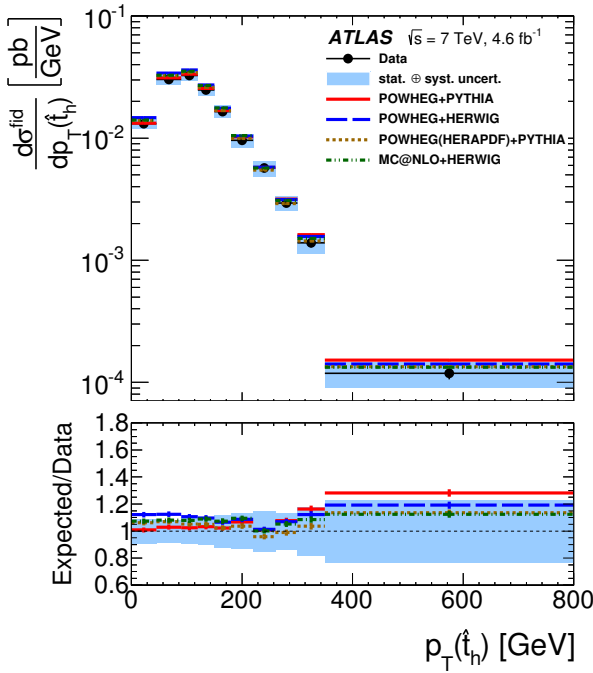
Figure 39: The small number of top jets identified at high-p_T results in very coarse cross section measurement when using the *resolved* reconstruction technique. Adapted from [169].

Figure 40: Particle-level differential $t\bar{t}$ cross section measurement for two different PDF sets and choices of the h_{damp} parameters. Taken from Ref. [173].
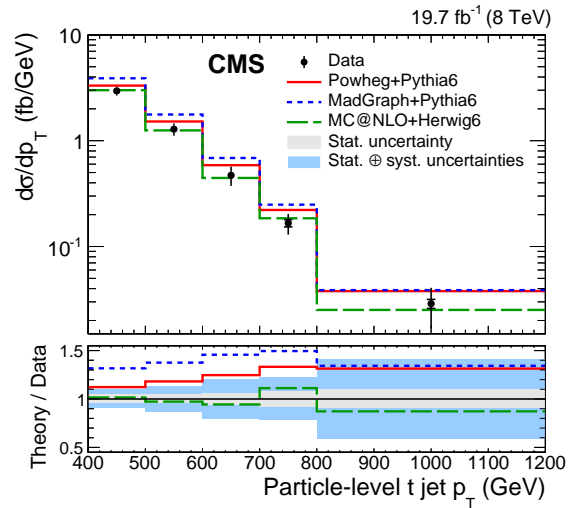
Figure 41: Comparison of the particle-level differential $t\bar{t}$ cross section as a function of the jet p_T to three different MC generators. Taken from Ref. [174].

The parton-level differential cross section in top quark p_T has also been measured in the all-hadronic final state by the CMS Collaboration using 8 TeV data [175]. This measurement relies on pruned jets with an N-subjettiness and subjet-b tagging requirement to suppress the huge amount of background from QCD dijet production. The cross section is determined from a maximum likelihood fit to the jet mass distributions for signal-enriched and signal depleted regions. This allows for a simultaneous extraction of the $t\bar{t}$ cross section while constraining the QCD background. The measurement is in agreement with the results from the lepton+jets final states, but has somewhat larger statistical uncertainties of up to about 40% in the highest p_T bin with $0.8 < p_T < 1.2$ TeV.

The increased $\sqrt{s}$ at Run 2 of the LHC offers the possibility for more precise differential $t\bar{t}$ cross section measurements in the highly-boosted regime. The $t\bar{t}$ production cross section increased by more than a factor of ten for top quark $p_T > 400$ GeV when going from $\sqrt{s} = 8$ TeV to 13 TeV.

A first measurement based on 3.2 fb^{-1} of 13 TeV data

in the lepton+jets channel has been performed by AT-LAS [172]. The measurement extends to p_T of 1.5 TeV and a similar trend as at 8 TeV is observed between the data and the simulation at high p_T. A newer measurement of the $t\bar{t}$ differential cross section in the all-hadronic channel is performed by the ATLAS Collaboration with 36.1 fb^{-1} of 13 TeV data [176]. The measurement uses trimmed anti-k_t $R = 1.0$ jets with $R_{\mathrm{sub}} = 0.2$ and $f_{\mathrm{cut}} = 0.05$. To obtain a flat signal efficiency of 50%
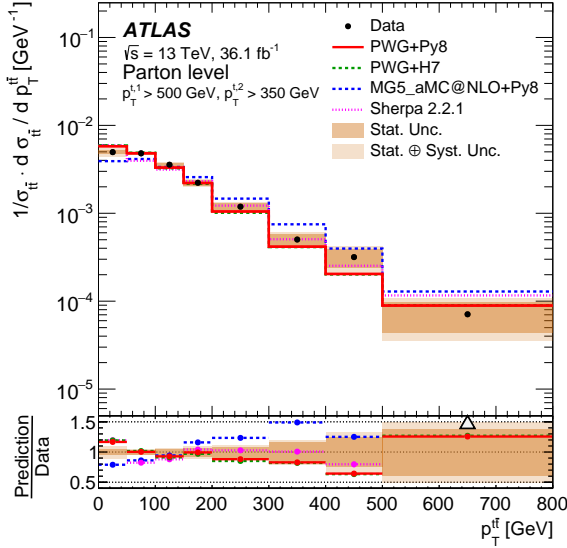
Figure 42: The normalized differential cross section as a function of the $t\bar{t}$ p_T as measured by ATLAS in the all-hadronic channel at 13 TeV. Taken from Ref. [176].

and a quark/gluon rejection of approximately 17 (10) for $p_T = 500$ (1000) GeV, p_T dependent criteria are applied on the jet mass and τ_{32}. Furthermore the two top-tagged large-R jets are required to have a b-tagged small-R jet within $\Delta R < 1.0$. The event selection results in a signal-to-background ratio of 3-to-1. The measured fiducial phase-space cross section is $\sigma = 292 \pm 7$ (stat) ± 76 (sys) fb compared to the Powheg+Pythia8 prediction of 384 ± 36 fb at NNLO+NNLL. The measured normalized differential cross section as a function of the top jet p_T and rapidity are in good agreement with the different MC predictions. Larger discrepancies are observed for the p_T of the $t\bar{t}$ system as shown in Fig. 42. The measurement is dominated by the systematic uncertainties on the jet energy, mass and substructure scale of the large-R jets, alternative parton shower model and the uncertainties on the b-jet identification.

4.1.2 W/Z/H Cross Sections

The cross section of boosted W and Z bosons was measured by ATLAS in 4.6 fb^{-1} of 7 TeV pp collisions [43]. The hadronically decaying W and Z bosons are reconstructed as one single ungroomed anti-k_t $R = 0.6$ jet with $p_T > 320$ GeV, $|\eta| < 1.9$ and masses ranging between 50 and 140 GeV. The W and Z signal is en-

hanced over the dominating QCD background by constructing a likelihood discriminant from three substructure variables; thrust minor [177, 178], sphericity [179] and aplanarity [179], resulting in a signal efficiency of 56% and a background rejection of 89%. The jet mass distribution after subtracting the expected background from $t\bar{t}$ events is shown in Fig. 43. A binned maximum likelihood fit to the jet mass distribution is used to extract the W/Z jet signal yield and to calculate the inclusive cross section. Only the combined $W + Z$ cross section measurement is performed in this analysis due to the limited jet mass resolution. The combined $W + Z$ cross section is measured to be $\sigma_{W+Z} = 8.5 \pm 0.8$ (stat.) ± 1.5 (syst.) pb and is in agreement with the Standard Model prediction of $\sigma_{W+Z} = 5.1 \pm 0.5$ pb within 2 standard deviations. The dominating systematic uncertainties are the jet mass resolution and the choice of the QCD background pdf. The signal significance was furthermore studied when using groomed jets instead of ungroomed jets. Without an optimization of the analysis for groomed jets, similar significances were observed for groomed and ungroomed jets as expected due to the low number of pileup vertices in the 7 TeV dataset.
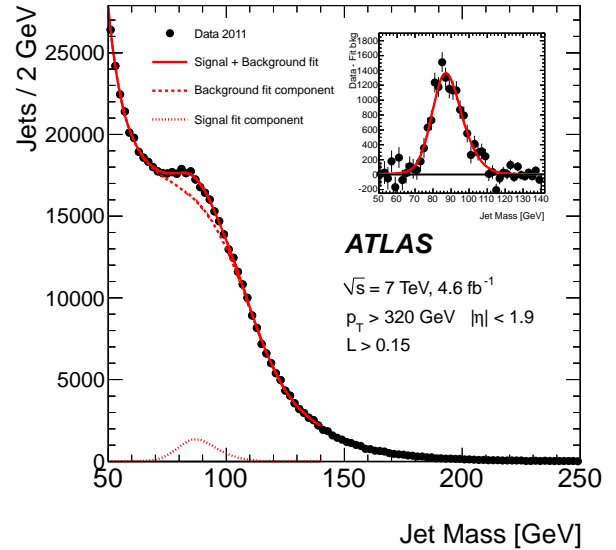

Figure 43: Binned maximum likelihood fit to the jet mass distribution in data for selected W/Z events reconstructed as one single ungroomed anti-k_t $R = 0.6$ jet. Taken from Ref. [43].

The ATLAS Collaboration also measured the high p_T

$Z \rightarrow b\bar{b}$ cross section using two b-tagged anti-k_t $R = 0.4$ jets in 19.5 fb^{-1} of 8 TeV pp collisions instead of one single large-R jet [180]. The measured fiducial cross section was determined to be $\sigma_{Z \rightarrow b\bar{b}} = 2.02 \pm 0.33$ pb which is in excellent agreement with the next-to-leading-order theoretical predictions.

As discussed in Section 3.4 the SM Higgs boson decays with approximately 58% into $b\bar{b}$. However the $H \rightarrow b\bar{b}$ decay in the resolved channel can only be studied in the associated production with either a vector boson (W/Z) [181, 182] or top quark or via the vector-boson-fusion production mechanism due to the overwhelming multijet background. To search directly for $H \rightarrow b\bar{b}$, jet substructure techniques can be employed to suppress the enormous multijet background. The CMS Collaboration performed a search for the SM Higgs boson using a single-jet topology with 35.9 fb^{-1} of 13 TeV pp collisions [183]. The analysis uses anti-k_t $R = 0.8$ jets that are modified with the soft drop algorithm ($\beta = 0$, $z_{\text{cut}} = 0.1$) to mitigate the effects from pileup and to remove wide-angle radiation. At least one large-R jet with $p_T > 450$ GeV and $|\eta| < 2.5$ is required. To distinguish the two prong structure of a jet containing the full $H \rightarrow b\bar{b}$ decay from quark- or gluon-initiated jets, the N_2^1 variable, calculated from the generalized energy correlation functions, is exploited. To ensure a flat QCD background rejection of 26% over the considered mass and p_T range, a decorrelation procedure [129] is applied to N_2^1. The multijet background is further suppressed by utilizing the double-b tagger which exploits not only the presence of two b quarks but also uses the fact that the flight direction of the b hadron is strongly correlated to the energy flow of the two subjets. The W/Z+jets background are estimated from MC simulation and the shape of the multijet background is determined in a validation region in data with lower values of the double-b tagger discriminator. The soft drop mass distribution of the leading jet is shown in Fig. 44 with a clear resonant structure at the mass of the W and Z boson. The SM background processes and the potential signal from SM $H \rightarrow b\bar{b}$ production are estimated simultaneously. The measured cross section for the Z+jets process is 0.85 ± 0.16 (stat.) $^{+1.0}_{-0.4}$ (syst.) pb which is in agreement with the SM prediction of 1.09 ± 0.11 pb. This is the first observation of $Z \rightarrow b\bar{b}$ in the single jet topology. The observed (expected) significance for the $H \rightarrow b\bar{b}$ process is $1.5(0.7)\sigma$.
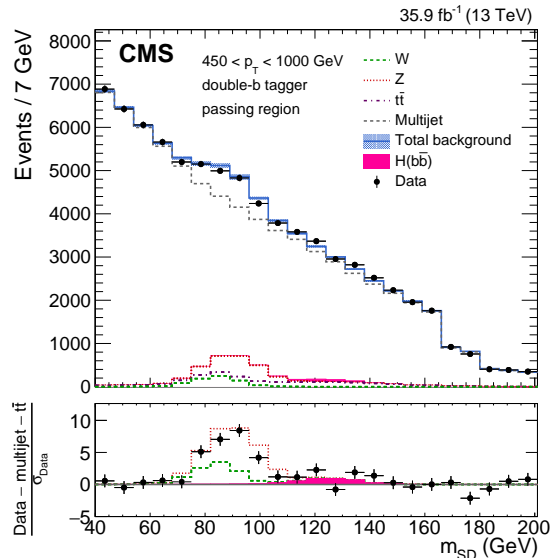
Figure 44: Soft drop jet mass m$_{\text{SD}}$ of anti-k_t $R = 0.8$ jets in data and for the dominating background processes; multijet production and W/Z+jets events. Jets are required to pass criteria on N_2^1 and to be identified as double-b jets by the double-b tagger introduced in Section 3.4. Taken from Ref. [183].

4.2 Measurements of jet substructure

The ATLAS and CMS Collaborations have performed a series of precision measurements of hadronic jet substructure in pp collisions, correcting for acceptance and resolution such as jet and event shapes [184, 185, 186, 187], charged particle multiplicities [187, 188, 189], the jet fragmentation functions [190], color flow [191] and k_T splitting scales, N-subjettiness ratios as well as further substructure variables such as Planar Flow and angularity [192, 193, 193]. This section will focus on measurements of the jet mass and the jet charge [194, 195].

4.2.1 Jet mass

The particle-level $t\bar{t}$ cross section was measured differentially as a function of the top-jet mass by the CMS Collaboration in lepton+jets events using 8 TeV data [67]. Large-R jets are reconstructed with the C/A algorithm using a distance parameter of 1.2 and $|\eta| < 2.5$. The larger value of R in this measurement compared to the default $R = 0.8$ applied for top tagging applications in CMS is due to an optimization of statistical precision due to the

Figure 45: Particle-level differential $t\bar{t}$ cross section measurement as a function of the leading jet mass compared to the predictions for three different MC generators. Taken from Ref. [67].

Figure 46: Normalized particle-level differential $t\bar{t}$ cross section measurement as a function of the leading jet mass compared to predictions using three different top quark mass values. Taken from Ref. [67].
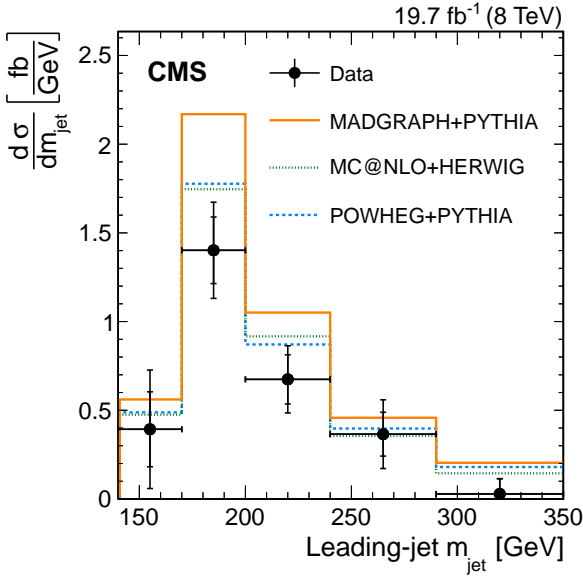
number fully-merged top quarks and the width of the jet mass distribution, which grows with increasing R. The leading jet p_T is required to be above $400\,\text{GeV}$ to ensure the hadronic top quark decay to be fully captured within the large-R jet. A requirement of $p_T > 150\,\text{GeV}$ is imposed on the subleading jet to select the b quark from the leptonically decaying top quark. The particle-level differential $t\bar{t}$ cross section as a function of the leading jet mass is shown in Fig. 45. The shown MC predictions exceed the measurement, consistent with the cross section measurements from the ATLAS and CMS Collaborations at high p_T. The total cross section in the fiducial region is $\sigma_{t\bar{t}} = 101 \pm 11\,(\text{stat}) \pm 13\,(\text{sys}) \pm 9\,(\text{model})$ for $140 < M_\text{jet} < 350\,\text{GeV}$. The experimental systematic uncertainties are dominated by the uncertainties on the jet mass and energy scale, but are smaller than the uncertainties due to the signal modeling, coming from the choice of the top quark mass, the parton showering and the choice of the factorization and renormalization scales.

For the first time, the top quark mass was determined in boosted $t\bar{t}$ events by the CMS Collaboration. The top quark mass was extracted from the normalized differential cross section measurement, shown in Fig. 46, since only the shape of the top-jet mass distribution can be re-

liably calculated. The top quark is measured to be $m_t = 170.8 \pm 6.0\,(\text{stat}) \pm 2.8\,(\text{sys}) \pm 4.6\,(\text{model}) \pm 4.0\,(\text{theo})\,\text{GeV}$ which is in agreement with top quark mass measurement in resolved $t\bar{t}$ events (see e.g. Refs. [196, 197]). The usage of boosted techniques is of particular theoretical interest to perform precision measurements of the top quark mass. The most precise top quark mass measurements [196, 198, 199] are currently based on template and matrix element fits to the reconstructed top quark decay products and thus measure the Monte Carlo top quark mass m_t^{MC} which is often assumed to be the same as the pole mass but depends in fact on the parton shower and hadronization model. The systematic uncertainties of these measurements are dominated by MC modeling uncertainties. The reconstruction of the top quark decay products as one single jet allows for the extraction of the top quark mass without ambiguities that arise from the translation of the MC mass into a short-distance mass scheme [200]. Furthermore it has been shown that the usage of grooming techniques can significantly reduce the MC modeling systematic uncertainties [201].

The first measurement of the normalized dijet differential cross section as a function of the jet mass was performed by the ATLAS Collaboration with a dataset corresponding to $35\,\text{pb}^{-1}$ of $7\,\text{TeV}$ pp collisions [192]. Both the

cross section for groomed and ungroomed C/A $R = 1.2$ jets was measured separately to gain sensitivity to both the hard and soft jet physics and to gain a deeper understanding of the various effects involved in QCD radiation. For the ungroomed jet mass, large discrepancies were observed in the tails of the mass distribution between the MC predictions from Pythia and Herwig++ and the data whereas the core of the mass distribution agreed within approximately 20% over the considered p_T range. The largest discrepancies occur at low jet masses which is sensitive to the underlying event description, hadronization model and pileup effects. The normalized cross section after applying the split filtering algorithm with $\mu_{\text{frac}} = 0.67$ and $y_{\text{filt}} = 0.09$ is shown in Fig. 47. After removing soft radiation from the jet which is difficult to model, the MC prediction is in excellent agreement with the data within statistical precision. The CMS Collaboration performed a similar measurement with anti-k_t $R = 0.7$ jets using various grooming techniques in selected dijet events using 5 fb^{-1} of $\sqrt{s} = 7$ TeV data and found as well that the agreement between data and the MC prediction improves significantly after grooming techniques are applied [53]. Furthermore a measurement of the cross section was performed in $V+$jet final states which overall show a slightly better data/MC agreement than that observed in dijet events suggesting that the simulation of quark jets is better than for gluon jets.

The CMS (ATLAS) Collaboration measured the double-differential jet cross section in balanced dijet events at $\sqrt{s} = 13$ TeV for groomed anti-k_t $R = 0.8$ jets with the soft drop algorithm with $z_{\text{cut}} = 0.1$ and $\beta = 0$ ($\beta = 0, 1, 2$) [42, 54]. The soft drop algorithm was chosen as it allows to compare the unfolded measurement directly to theoretical calculations which significantly exceed the precision of parton shower MC simulations. The jet energy of the ungroomed jets used in the ATLAS measurement are corrected for pileup effects and calibrated to the generator-level while no explicit mass calibration is applied to the groomed jets as the unfolding procedure accounts for differences between the reconstructed and generator-level mass. The CMS Collaboration applied calibration factors derived from simulation and using *in-situ* techniques to correct the jet energy and mass scale. Furthermore the jet energy and mass are smeared in MC simulation to match the resolution measurements in data. Various sources of systematic uncertainties, categorized as experimental and theoretical uncertainties, that impact the jet mass measurement are taken into account.
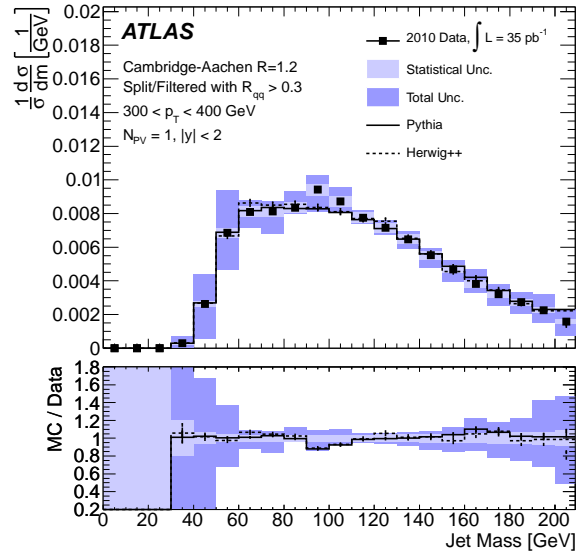
Figure 47: Normalized differential cross section as a function of the jet mass for C/A jets with $R = 1.2$ after splitting and filtering, taken from Ref. [192].

While CMS evaluated the effect of the jet energy and mass scale uncertainties on the measurement by varying the energy and mass by their respective uncertainties, ATLAS evaluated the experimental uncertainties based on the accuracy of the modelling of the topological cluster energies and positions as well their reconstruction efficiency. Theoretical uncertainties on the physics model are taken into account by comparing the response matrix for various MC generators.

The comparison of the normalized cross section with two analytical calculations as measured by CMS is shown in Fig. 48. ATLAS measured instead the unfolded $\log_{10} \rho^2$ distribution, shown in Fig. 49, where ρ is the ratio of the soft drop jet mass to the ungroomed jet p_T. Both measurements are compared to calculations at next-to-leading order with next-to-leading-logarithm and leading order with next-to-next-to-leading-logarithm accuracy. Good agreement between the data and the predictions is observed in resummation regime $-3.7 < \log_{10} \rho^2 < -1.7$. For higher jet masses, where non-perturbative effects play an important role, the NLO+NLL calculation, which includes non-perturbative effects, provides a better description than the LO+NNLL calculation.
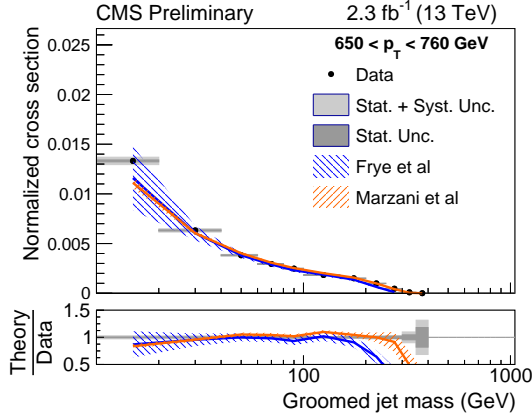
33

Figure 48: Normalized differential cross section as a function of the mass for jets groomed with the soft drop algorithm in data and for two theoretical calculations. Taken from Ref. [54].
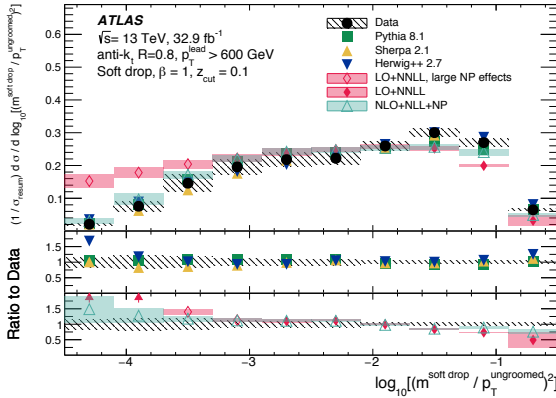

Figure 49: Comparison of the unfolded $\log_{10} \rho^2$ distribution for $z_{\text{cut}} = 0.1$, $\beta = 1$ in data to various Monte Carlo particle-level predictions and theory predictions, normalized to the integrated cross section measured in the resummation regime $-3.7 < \log_{10} \rho^2 < -1.7$. Taken from Ref. [42].

4.2.2 Jet Charge

This section summarizes the measurements of the jet charge [194, 195], defined as the energy weighted sum of the electric charges of the jet constituents

$$Q_\kappa = \sum_{i \in J} \left(\frac{p_{\text{T},i}}{p_{\text{T},J}} \right)^\kappa q_i \,, \tag{5}$$

where q_i is the electric charge of particle i and the free parameter κ that controls the sensitivity to soft particles within the jet. The ATLAS (CMS) Collaboration measured the jet charge for different values of κ using anti-k_t jets with a radius parameter of $R = 0.4$ ($R = 0.5$) in a sample of dijet events. The ATLAS Collaboration distinguishes between the two leading jets using the pseudorapidity instead of the p_{T} to avoid cases where the leading particle-level jet is reconstructed as the sub-leading detector-level jet due to the jet energy resolution. The average jet charge at detector- and particle-level for the more forward of the leading jets and for $\kappa = 0.5$ is shown in Fig. 50. Due to the increasing fraction of scattering valence up quark jets (up quark charge > 0), the average jet charge increases with p_{T}. The difference of the average jet charge distribution at detector-level and particle-level in Fig. 50 shows that the unfolding corrections are large and growing at high p_{T}, due to the loss of charged-particle tracks inside jets as a result of track merging. The average jet charge as predicted by Pythia 8 [202] using the Perugia tunes [203] is smaller than that observed in data due to a well-known over-estimation of the multiplicity inside jets. The dominating systematic uncertainties are the track p_{T} resolution and the choice of MC generator used to construct the response matrix (Pythia 6 versus Herwig++) for the CMS Collaboration whereas the uncertainties on the unfolding procedure, the jet energy resolution at low p_{T} and uncertainties on the tracking at high p_{T} dominate the measurement of the ATLAS Collaboration. The unfolded jet charge distribution ($\kappa = 0.6$) of the leading jet in data is compared to the prediction from Powheg+Pythia8 (PH+P8) and Powheg+Herwig++ (PH+HPP) in Fig. 51. The different hadronization and fragmentation model used by Pythia8 and Herwig++ have the largest impact on the jet charge distribution. Variations of the jet charge can also be observed for different PDF sets however the effect of the relative flavor fraction in the dijet samples is significantly smaller than the choice of the showering and fragmentation model. It was further found that the predicted jet charge distribution has a significant dependency on the chosen value of α_s that describes final state radiation whereas it is insensitive to NLO QCD effect in the matrix element calculation, color-reconnection and multiple parton interactions. These findings are consistent between the ATLAS and CMS Collaboration.

In addition to studying the sensitivity to various non-perturbative aspects of hadronization and parton distribution functions, the jet charge measurement by ATLAS

includes the first direct comparison of a jet substructure quantity with a perturbative calculation at a hadron collider. As it is not collinear safe, the average jet charge is not calculable. However, the p_T dependence for a particular jet type has been calculated [204, 205]. A new technique was introduced in Ref. [194] to separately extract the average up and down quark jet charge. For a fixed p_T, the more forward of the two dijets has a higher energy and is therefore more likely to be the scattering parton with a higher momentum fraction of the proton. In turn, the higher momentum fraction parton is most likely to be a valence quark. Therefore, the fraction of up quark jets is higher for the more forward dijet than the more central dijet. Assuming further that the jet charge is entirely determined by the jet p_T and parton origin, one can then solve a system of equations to extract the average up and down quark jet charge in each bin of jet p_T:

$$\langle Q_J^f \rangle = f_u^f \langle Q_J^u \rangle + f_d^f \langle Q_J^d \rangle$$
$$\langle Q_J^c \rangle = f_u^c \langle Q_J^u \rangle + f_d^c \langle Q_J^d \rangle, \tag{6}$$

where f = forward, c = central, u = up and d = down. As expected (though not an input), the average up quark charge is positive and the average down quark charge is negative; furthermore, the latter is roughly half the former in absolute value. The p_T dependence of $\langle Q_J^{u,d} \rangle$ are fit with a logarithmic scale violating term c: $\langle Q \rangle_i = \langle Q \rangle_0 (1 + c_\kappa \ln(p_{T,i}/p_{T,0}))$, where i represents the p_T bin. Figure 52 shows the measured and predicted values of c_κ. The uncertainties are large, but there is an indication that $c < 0$ and $\partial c / \partial \kappa < 0$, as predicted.

5 Searches for New Physics

Jet substructure methods have been successfully applied in a large variety of searches for physics beyond the SM. The respective exclusion limits are substantially improved through the application of these methods. In some cases the decay signature of heavy BSM particles would not be accessible without the application of jet substructure methods.

As the number of such BSM searches is very large, only a small subset of the published results can be discussed here. The following sections give an overview of a selection of searches for $t\bar{t}$ resonances [150, 206, 207, 208, 209, 210, 211], diboson resonances [66, 123, 124, 125, 126, 127, 212, 213, 214, 215, 216, 217, 218,
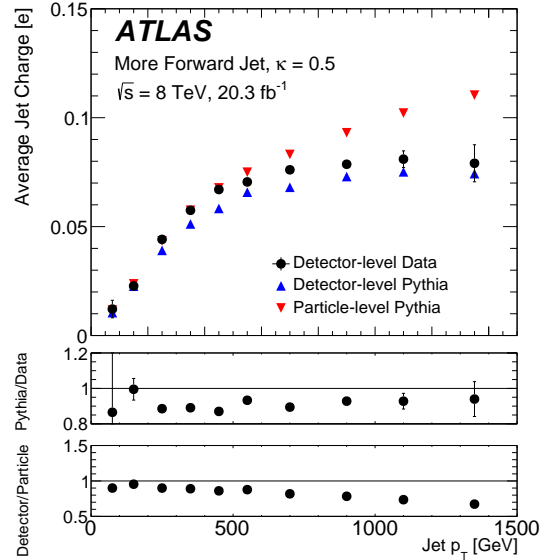
Figure 50: The detector- and particle-level average jet charge as a function of jet p_T. Reproduced from Ref. [194].
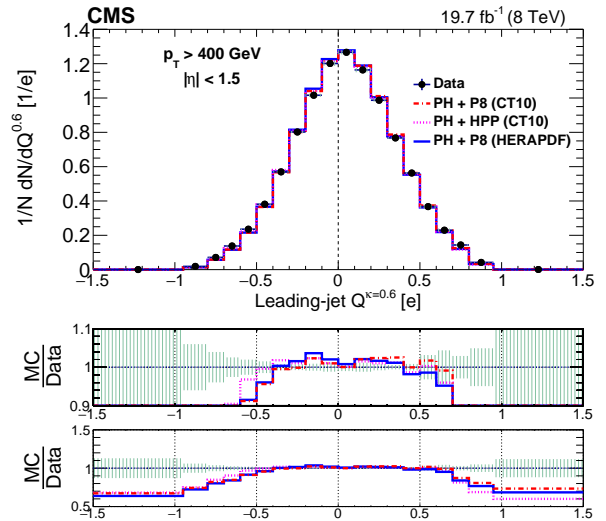
Figure 51: Unfolded jet charge distribution for $\kappa = 0.6$ in data and MC prediction. Taken from Ref. [195] .

219, 220, 221, 222, 223, 224, 225, 226, 227], vector-like quarks [228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243] and leptophobic Z' [122, 244]. Further searches using jet substructure techniques can be found in Refs. [47, 106, 145, 151, 245, 246, 247, 248, 249,
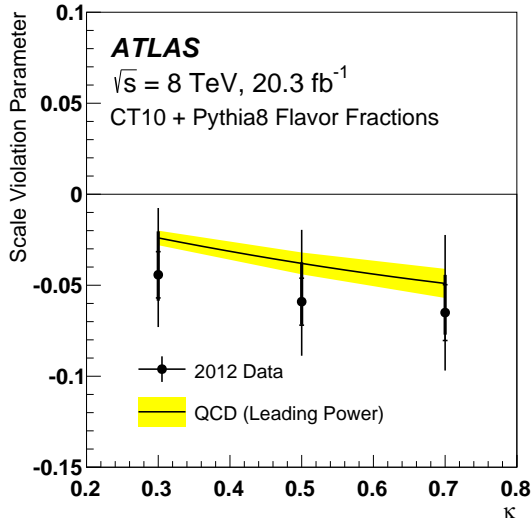
Figure 52: The measured and predicted value of the average jet charge scale violation parameter c_κ. Reproduced from Ref. [194].

250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270].

5.1 $t\bar{t}$ Resonances

Several models of new physics predict resonances coupling strongly to the top quark to play a role in the cancellation of large corrections to the Higgs mass. These models include extensions of the SM Higgs doublet, where the simplest realizations are two-Higgs-doublet models [271] with heavy, neutral Higgs bosons, which can have large branching fractions to top quarks. Alternative mechanisms of electroweak symmetry breaking often include new heavy particles with large couplings to top quarks. Examples for such models are the topcolor model containing a Z' boson [272] as well as composite Higgs models [273, 274, 275, 276, 277, 278, 279, 280]. The Randall-Sundrum Kaluza-Klein model predicts a massive gluon decaying to top quarks [281, 282, 283, 284]. While small resonance masses are mostly excluded, the searches are generally focusing at high masses with $m_{t\bar{t}} > 1$ TeV such that the top quarks receive high Lorentz boosts and must be reconstructed with jet substructure methods. In the following sections the analysis strategies and results from CMS and ATLAS using proton-proton collisions at 8 TeV are discussed, followed by a summary of the first results with 13 TeV data.

Lepton+jets Channel Due to the high boost of the top quarks from a heavy resonance decay, the lepton from the W boson decay may overlap with the associated b quark jet. Therefore, the lepton-isolation criteria, which are used to mitigate the contamination with QCD multijet background, are relaxed. The CMS and ATLAS Collaborations follow different strategies for this purpose. In the CMS analysis [207], the lepton must have a large angular separation from the associated b-jet candidate of $\Delta R(\text{lepton}, \text{jet}) > 0.5$ or it must have a transverse momentum relative to the jet axis p_T^{rel} above 25 GeV. This selection requirement removes background contributions from semi-leptonic B hadron decays. In the ATLAS analysis [150], the lepton isolation is achieved by a so-called mini-isolation requirement which applies a variable isolation cone that changes as a function of the lepton energy [285]. Interestingly, studies performed in CMS for 13 TeV show that the CMS implementation of mini-isolation is not as powerful as their own method for $t\bar{t}$ resonance searches.

To reconstruct the boosted hadronic top decay, the presence of a single high-momentum, large-R, top-tagged jet is required. In CMS (ATLAS) the large-R jet is reconstructed with the Cambridge/Aachen (anti-k_t) algorithm with a size parameter of $R = 0.8$ (1.0). The selection requirement on the transverse momentum is $p_T > 400(300)$ GeV. ATLAS applies trimming to the large-R jets with the parameters $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$ and the jets are required to have a mass $m_{\text{jet}} > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV. The strategy followed by CMS is to apply the "CMS top-tagging" algorithm (as defined in Section 3.3), where the mass of the jet has to satisfy $140 < m_{\text{jet}} < 250$ GeV. In addition, the N-subjettiness ratio τ_{32} must be smaller than 0.7.

The variable of interest is the invariant mass $m_{t\bar{t}}$ of the $t\bar{t}$ system. It is reconstructed from the top-tagged large-R jet, a b-tagged small-R jet as well as the lepton and the missing energy. Once the top pair system is reconstructed, events are further divided into categories based on the lepton flavor and the number of b-tagged and top-tagged jets. This gives several analysis categories with different background compositions: the top-tagged and b-tagged events are dominated by the SM $t\bar{t}$ background, while events without top-tags and b-tags are mostly composed of W+jets events. No excess of data events is seen over expected SM backgrounds, and exclusion limits are

set on the Z' and RS-KK gluon physics models, which are discussed in more detail below.

Hadronic channel An analogous search is performed by CMS assuming hadronic decays for both top quarks [207]. The same CMS top-tagging algorithm as in the lepton+jets channel is applied. In addition, the region with jets of lower momenta with $200 < p_T < 400\,\mathrm{GeV}$ is covered with a dedicated algorithm using a larger jet size parameter of $R = 1.5$ (CA15 jets). The larger jet size allows to cover topologies with larger angles between the decay products of the top quark in case of intermediate or smaller Lorentz boosts. These low-p_T jets are required to be identified as top-jets by the HEPTopTagger algorithm (as described in Section 3.3). This approach allows to improve the sensitivity for smaller masses of the hypothetical $t\bar{t}$ resonance.

Additionally, events are divided based on the rapidity separation $|\Delta y|$ of the two jets. Jets from a high-mass particle will in general have a lower $|\Delta y|$ compared to QCD dijet events, where the high mass usually comes from a larger rapidity separation. Two regions are used to divide events, defined by $|\Delta y|$ greater or less than 1.0. This results in six event categories for the hadronic analysis.

Even with the requirement of two top-tagged jets, the selected sample is dominated by QCD dijet events. This background is estimated using a data-driven technique, where an anti-tag and probe method is used. The τ_{32} requirement is reversed on one jet to select a sample dominated by QCD events. The opposite jet is then used to measure the misidentification rate for the top-tagging requirements. The measured misidentification rate ranges from 5 to 10%, depending on the jet momentum, τ_{32} and the b-tagging requirements applied. This differential rate is used in a sample of single top-tagged events to predict the double top-tagged event contribution from QCD processes in each individual event category. Closure tests performed in data and simulation are performed to validate the background estimation for each of the signal regions.

No significant excess above the predicted background is observed in the measured $t\bar{t}$ invariant mass spectrum. The exclusion limits obtained on the Z' and RS-KK gluon models are discussed below.

Results The analysis of the ATLAS Collaboration [150] finds upper limits on the cross section times branching ratio of a narrow Z boson decaying to top quark pairs in the range from 4.2 pb to 0.03 pb for resonance masses
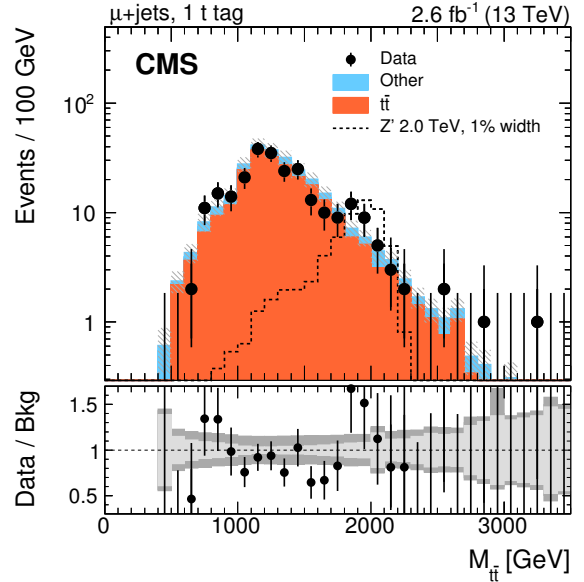
Figure 53: Invariant mass of the reconstructed $t\bar{t}$-pair in data and simulation for the lepton+jets channel in the category with one top-tagged jet, taken from Ref. [211].

ranging from 0.4 TeV to 3.0 TeV. The limits for larger widths (15%) are slightly worse.

The CMS Collaboration performs a combination of the lepton+jets, hadronic and di-leptonic channels. The obtained limits on the production cross sections range from 0.86 pb to 0.0059 pb for narrow resonances with masses from 0.75 TeV to 3.0 TeV, excluding a topcolor narrow Z' resonance with a mass below 2.4 TeV. The limit for a Z' boson with a width of 10% is 2.9 TeV, while Kaluza-Klein gluon excitations predicted in Randall-Sundrum models are excluded below 2.8 TeV.

First results with 13 TeV data A similar analysis strategy as for 8 TeV is adopted by the ATLAS and CMS Collaborations for the first data at the LHC collision energy of 13 TeV [210, 211]. Figure 53 shows the $m_{t\bar{t}}$ spectrum in the category with the highest S/B fraction in the CMS analysis which does not exhibit any deviations from the SM background. With an integrated luminosity of 3.2 and 2.6 fb^{-1}, Topcolor Z' bosons with relative widths of 1%, 10%, and 30% are excluded for the mass ranges of 0.6–2.5, 0.5–3.9, and 0.5–4.0 TeV, respectively. Kaluza-Klein gluon excitations predicted in Randall-Sundrum models are excluded below 3.3 TeV. These limits are significantly more stringent than those obtained with 8 TeV data.

5.2 Diboson Resonances

The new physics models discussed in the previous section also predict new resonances decaying to pairs of W, Z or Higgs bosons. In more than 60% of the cases, $W/Z/H$ bosons decay into a quark anti-quark pair, which makes the reconstruction of such decays an essential ingredient for these searches.

The search for diboson resonances is performed at $\sqrt{s} = 13\,\mathrm{TeV}$ in semi-leptonic [220, 225?] and fully hadronic final states [66, 125, 126, 127] by the ATLAS and CMS Collaborations. As the methods of jet substructure analyses exhibit their full strength in hadronic final states, the following discussion gives a summary and comparison of the ATLAS and CMS results in the search for W/Z resonances in hadronic final states only.

In the analysis performed by the CMS Collaboration [66] events with two anti-k_t jets with $R = 0.8$ and $65 < m_{\mathrm{soft\ drop}} < 105\,\mathrm{GeV}$ are selected, where the jet is called a W boson candidate if the mass is in the range 65–85 GeV, while it is called a Z boson candidate if the mass is in the range 85–105 GeV. This leads to the three signal categories WW, ZZ and WZ. The jets are further categorized according to τ_{21} into high purity (HP, $\tau_{21} < 0.35$) and low purity (LP, $0.35 < \tau_{21} < 0.75$). Events are always required to have one HP V jet, and are divided into HP and LP events, depending on whether the other V jet is of high or low purity. To further suppress the large QCD multijet background a requirement on the dijet kinematics $|\eta_1 - \eta_2| < 1.3$ is applied.

The background is estimated from a signal + background fit with the function $\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_1}}$, where P_0 is a normalization parameter and P_1 is a parameter describing the shape. As shown in Fig. 54 the data is well described by the fit function. Figure 54 also shows that no excess over the background-only hypothesis is observed. Exclusion limits are set on new physics resonances under different hypotheses of spin and charge. W' and Z' resonances with masses below 3.6 and 2.7 TeV are excluded for the heavy vector triplet model B [286].

A similar analysis has been performed by the ATLAS Collaboration [127]. In this analysis events are required to have at least two large-R jets with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ in the pseudo-rapidity range $|\eta| < 2.0$. These jets are reconstructed with the anti-k_t algorithm with a radius parameter $R = 1.0$. The trimming algorithm is applied using k_{T} subjets with $R = 0.2$. The rapidity separation between the two leading jets has to satisfy $|\Delta y_{12}| < 1.2$.
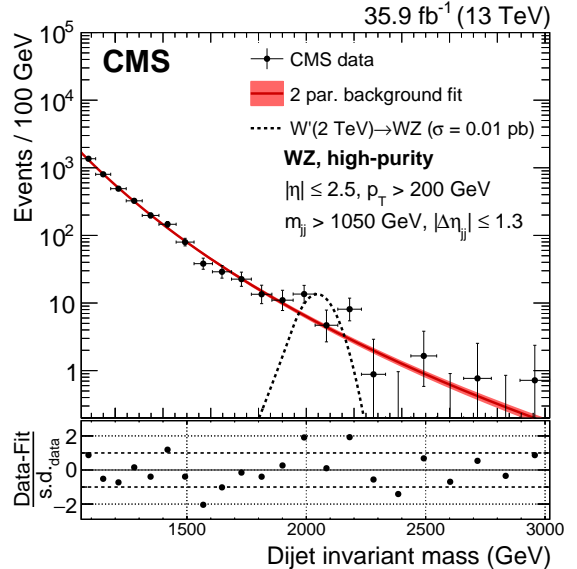
Figure 54: Dijet invariant mass distribution in the high purity WZ category of the fully hadronic $WW/WZ/ZZ$ resonance search. The fit under the background-only hypothesis is overlaid. Taken from Ref. [66].

The large-R jet mass is computed from the Combined Mass (see Sec. 1.5), and is required to be within a window of the expected W or Z mass value. The window width varies from 22 to 40 GeV depending on the jet p_{T}. In addition, the $D_2^{\beta=1}$ variable is used to select jets with a two-prong structure.

Similar as in CMS analysis, the background is estimated by fitting the dijet invariant mass distribution with the parametric form $\frac{dn}{dx} = p_1(1-x)^{p2+\xi p_3}x^{p3}$, where n is the number of events, x is a dimensionless variable related to the dijet mass m_{JJ}, $x = m_{\mathrm{JJ}}/\sqrt{s}$, p_1 is a normalization factor, p_2 and p_3 are dimensionless shape parameters, and ξ is a constant chosen to remove the correlation between p_2 and p_3 in the fit.

The dijet invariant mass distributions for these events are shown in Figs. 55, where good agreement is found between data and the expectations from the background fit.

Limits are set on the cross section times branching ratio for diboson resonances. In the heavy vector triplet model B a spin-1 vector triplet is excluded for masses between 1.2 and 3.5 TeV.

The analyses described above investigate final states with boosted W and Z bosons. In case of boosted H

Figure 55: The observed data in the signal region of the WZ category. Also shown is the fitted background prediction. The gray region represents the uncertainty in the background estimate. Taken from Ref. [127].

Figure 56: Multiplicity of CA15 jets which fulfill the Higgs tagging criteria. The solid histograms represent the simulated background processes ($t\bar{t}$ and QCD multijet). The hatched error bands show the statistical uncertainty of the simulated events. Taken from Ref. [229].
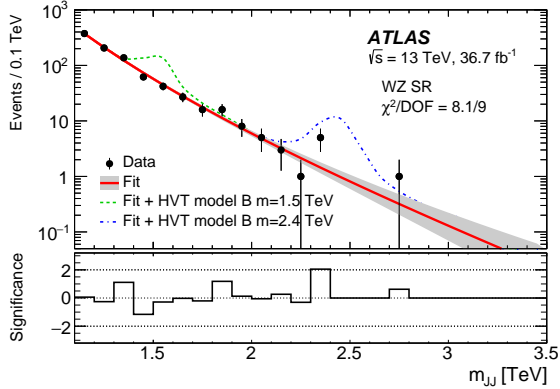
bosons, different reconstruction methods have to be used to benefit from the presence of b quarks in $H \rightarrow b\bar{b}$ decays (see Sec. 3.4). Results have been published on the search for for WH/ZH final states [126, 221, 226] as well as for HH final states [222, 223, 227] using proton-proton collisions at $\sqrt{s} = 13$ TeV.

5.3 Vector-like Quarks

Vector-like quarks (VLQs) are predicted by a variety of theories introducing a mechanism that stabilizes the mass of the Higgs particle. Such theories include little Higgs models [287, 288], models with extra dimensions [289, 290], and composite Higgs models [289, 290, 291]. As VLQs are expected to have large masses, jet substructure analyses have been applied in many searches for VLQs.

The first search for VLQs using jet substructure methods was an inclusive search for pair-produced T quarks [228]. This search was conducted by the CMS Collaboration using pp collisions at $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of 19.7 fb^{-1}. In this search the Cambridge/Aachen algorithm was used with a distance parameter $R = 0.8$ (CA8 jets). Boosted W jets are identified based on the mass of the CA8 jet while boosted top jets are identified with the CMSTT, described in Sec. 5.1. The limits on the hypothetical mass of a VLQ are between 687 and 782 GeV, depending on the decay modes.

The first search for VLQs in the all-hadronic final state has been performed by CMS and is described in Ref. [229].

The strategy of this analysis is to apply the C/A algorithm with a large size parameter of $R = 1.5$ to cluster the decay products from top quarks and Higgs bosons into single large jets. To identify the origin of the large C/A jets a top tagging algorithm (HEPTopTagger) and a Higgs tagging algorithm are used. This was the first time the HEP-TopTagger algorithm has been applied in a data analysis by the CMS Collaboration. This was also the first time a Higgs tagging algorithm based on subjet b-tagging (see Sec. 3.4) has been used. Two subjets must be b-tagged and their invariant mass must be greater than 60 GeV to fulfill the Higgs tagging requirement. The multiplicity of these Higgs tags is shown in Fig. 56 which demonstrates that both the QCD multijet and the $t\bar{t}$ backgrounds can be suppressed by several orders of magnitudes. The efficiencies of these algorithms were also measured, since these are needed for modelling the signal efficiencies. As a result of this analysis VLQ masses of less than 745 GeV have been excluded at 95% confidence level, assuming a 100% branching fraction into the tH final state.

Extensive use of substructure methods has also been made by the ATLAS Collaboration, in particular for the search for single production of VLQs. The single pro-

duction modes may have higher cross sections than pair production depending on the VLQ mass and the coupling parameters[292]. ATLAS performed an analysis [240] where the VLQ is searched for in the decay mode with a W boson and a top quark ($B \to tW$). Final states with at least one lepton are considered, where either the W boson or the top quark appear in a boosted configuration. They are identified by the application of a jet mass requirement ($m > 50\,\mathrm{GeV}$) on a trimmed large-R anti-k_t jet with a distance parameter $R = 1.0$.

A different strategy is followed in another ATLAS search [241], where the decay into the bW final state is investigated ($T/Y \to bW$). As the W boson is assumed to decay leptonically, no boosted hadronic W or top quark decays are present. Nevertheless, the analysis uses a veto on the presence of massive ($m > 70\,\mathrm{GeV}$), trimmed large-R anti-k_t jets with $R = 1.0$, to suppress the dominant $t\bar{t}$ background.

In the meantime jet substructure methods are widely employed in almost all VLQ searches at $\sqrt{s} = 13\,\mathrm{TeV}$ published by the LHC Collaborations, see e.g. Refs. [232, 233, 234, 235, 236, 237, 242]. Note that this list is not exhaustive by far and many more analyses are being finalized presently. The excluded VLQ masses are exceeding 1 TeV for all branching fractions. The sensitivity of these searches to VLQ masses up to and beyond 1 TeV is a huge success, which has been made possible thanks to jet substructure techniques.

5.4 Leptophobic Z'

Many extensions of the Standard Model predict resonances that couple to quarks and gluons [293, 294, 295, 296], including simplified Dark Matter (DM) models in which resonances couple only to quarks and DM particles [297, 298, 299].

The ATLAS and CMS Collaboration searched for low mass leptophobic Z' decaying into $q\bar{q}$ in the full 2015+2016 dataset collected at 13 TeV [122, 244]. The search for the Z' production is performed in association with a high p_T jet from initial-state radiation (or also a photon in the case of ATLAS) to be able to cover resonance masses of $m_{Z'} \ll 1\,\mathrm{TeV}$. The sensitivity to low resonance masses is usually degraded due to high trigger thresholds and the enormous QCD multijet background. The decay products of the Z' are collimated and captured within one single large-R jet. Jets in the CMS analysis are reconstructed with the anti-k_t algorithm with $R = 0.8$ and corrected for

pileup effects with the soft drop algorithm ($\beta = 0$, $z_\mathrm{cut} = 0.1$) whereas anti-k_t $R = 1.0$ jets, trimmed with $R_\mathrm{sub} = 0.2$ and $f_\mathrm{cut} = 5\%$ are used in ATLAS. To suppress the dominating QCD multijet background, CMS applies criteria on N_2^1 [115] and ATLAS chooses τ_{21} as discriminator. To avoid distortions of the jet mass spectrum due to large correlation between the jet mass and substructure variables, the substructure variables are decorrelated from the jet mass with the designed decorrelated tagger method. Data-driven techniques are used to determine the most-dominating backgrounds, namely the QCD multijet production whereas sub-domination processes such as W/Z+jets events are estimated from MC simulation. The jet mass distributions of the large-R jet is shown in Fig. 57 and 58 for the CMS and ATLAS analyses, respectively. No evidence for a resonant structure on top of the SM background is observed and the results are interpreted by CMS (ATLAS) as 95% confidence level limits on the production cross section and the coupling of the Z' to quarks, g_q, for Z' masses in the range from 50 to 300 GeV (100 to 220 GeV). The CMS Collaboration excludes coupling values of $g_q > 0.25$ over the full mass range, with stronger constraints for low jet masses. After combining the ISR jet and ISR photon channels, ATLAS sets upper limits on g_q of 0.17 for $m_{Z'} = 100\,\mathrm{GeV}$ and 0.21 for $m_{Z'} = 220\,\mathrm{GeV}$.
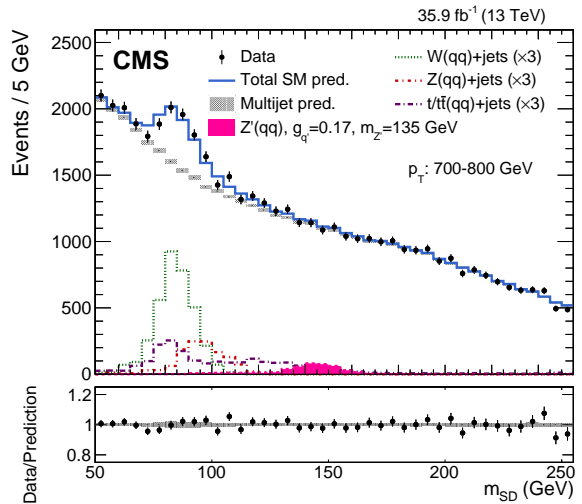

Figure 57: Soft drop jet mass of anti-k_t $R = 0.8$ jets in data and for the dominating background processes; multijet production and W/Z+jets events. Taken from Ref. [122].
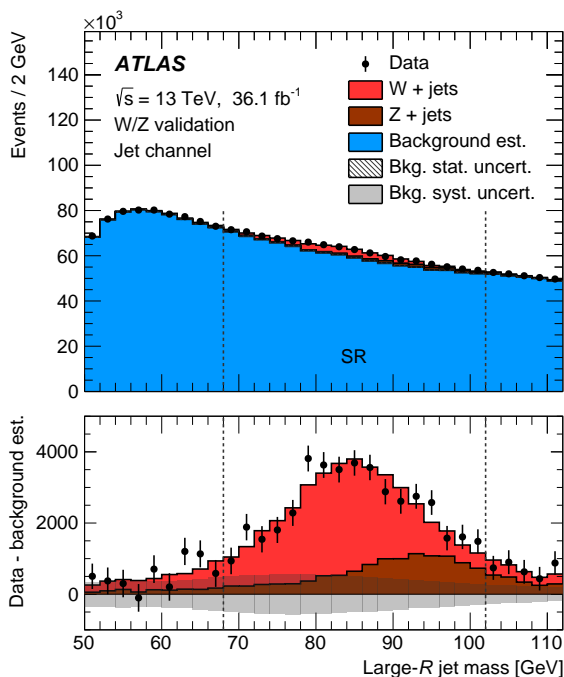
Figure 58: Trimmed jet mass distribution anti-k_t $R = 1.0$ jets in data and for the dominating background processes. Taken from Ref. [244].

6 Future Collider Projects

The LHC guarantees a rich scientific output in the next decade, stretching well into the 2030s with the planned luminosity upgrade. A diverse range of possible energy frontier machines beyond the LHC has been proposed. A summary is given in Table 3.

Lepton colliders come in two varieties: the linear collider projects promise scalable machines [300, 301, 302], that can explore high-energy processes such as vector-boson-fusion Higgs production, associated production of a top quark pair and a Higgs boson and di-Higgs boson production, and can ultimately reach the multi-TeV regime [303]. Large circular colliders (with a circumference of 100 km) can provide superior luminosity at the maximum of the ZH production cross section ($\sqrt{s} \sim 250$ GeV [304, 305]) and may reach the top pair production threshold [306].

Ambitious projects exist in Europe and China for a new pp collider with a size several times that of the LHC. The FCChh [308, 309, 310, 311] and SPPC [305] projects are set to reach a center-of-mass energy of 100 TeV. An inter-mediate step could be found in the upgrade of the LHC dipoles with 16 T bending magnets, which could roughly double the energy reach. Conceptual Design Reports are being prepared for the machines that have not yet presented detailed design, such that a complete assessment of scientific merits and financial cost can be made for example for the update of the European Strategy in 2019.

6.1 Hadron Colliders

A future high-energy hadron collider could reach an energy of up 100 TeV, with a luminosity that significantly exceeds that of the LHC [312]. Cross sections increase with center-of-mass energy. Combined, this yields a factor 100 increase in the sample of WW, ZZ, $t\bar{t}$ and Higgs events that can be collected. The increase is even larger for rare processes, such as $t\bar{t}H$ or di-Higgs boson production [313, 314] and is dramatic in the high-energy tails. A 100 TeV collider is a boosted-object factory, extending the mass reach for searches [309] to several tens of TeV. The study of boosted-object production rates also brings a much greater indirect sensitivity to new physics at even higher scale [315, 316, 317, 318].

With higher energies come new challenges. During LHC Run 1 and Run 2, a top quark is generally described as being boosted if it has p_T greater than twice its mass, e.g. $p_T > 350$ GeV, as this would indicate a high probability of capturing the hadronic decay products in a single jet with $R = 1.0$. In future ultra-high energy colliders, boosted object taggers must evolve to exploit different features for objects with p_T up to 10 or 20 TeV. At these energies, the decay products of multi-TeV top quarks, and especially W, Z and Higgs bosons can be contained by jets with $\Delta R < 0.1$.

The analysis of the substructure in jets formed by hyper-boosted objects pose stringent requirements on the granularity and size of the experiment. To resolve the substructure of jets at a very small scale the detector granularity and size must exceed that of current LHC hadron calorimeter significantly [319, 320]. To contain high-energy hadrons a deep hadronic calorimeter is required, with up to 12 interaction lengths [321] (cf. the CMS barrel calorimeter has a depth of 7 λ_{int}, including 1 λ_{int} in the EM system). The tracking efficiency in the core of jets, already under some pressure at the LHC, is another area where requirements related to jet substructure may condition the detector design. Detailed full-simulation studies are required to assess the experimental limitations.

Table 3: Projects for energy-frontier facilities. Integrated luminosities correspond to different running times (typically ten years per energy point) and are subject to large uncertainties. The last column provides a reference to the most recent design report. References to detailed running scenarios are given where available. All e^+e^- colliders envisage a brief period of running at energies close to the top quark pair production threshold, that is not listed here.

Energy-frontier lepton colliders					
Project	**Host**	**Type**	$\sqrt{s}$ (TeV)	$\int \mathcal{L}$ (ab^{-1})	**Status**
ILC	Japan	linear e^+e^-	0.25	0.5 (2)	TDR 2013 [301]
			0.5	0.5 (4)	staging [300, 307]
CLIC	CERN	linear e^+e^-	0.38	0.5	CDR 2012 [303]
			1.5,3	1.5,2	staging [302]
CEPC	China	circular e^+e^-	0.25	5	CDR 2017 [305]
FCCee	CERN	circular e^+e^-	0.25/0.36	10/2.6	CDR < 2019 [304]
μ collider	FNAL	racetrack $\mu^+\mu^-$	0.125-3	-	R&D

Very-high-energy pp colliders					
Project	**Host**	**Type**	$\sqrt{s}$ (TeV)	$\int \mathcal{L}$ (ab^{-1})	**Reference**
HE-LHC	CERN	LEP/LHC	25		-
FCChh	CERN	new	100	20	[308, 309, 310, 311]
SPPC	China	new	70-140	3	[305]

Several authors have explored different approaches to tag highly boosted top quarks. A "traditional" substructure analysis may be feasible [309] well into the multi-TeV regime. Larkoski et al. show that a substructure analysis limited to charged particle tracks provides significant separation power, thus avoiding strong constraints on the calorimeter granularity [322]. For top quarks, the lepton-in-jet signature [316] may be sufficient to isolate the signal.

Ref. [323] studies the prospect to identify highly boosted W, Z or Higgs bosons in 100 TeV collisions at the FC-Chh. As these are color singlets the energy of the boson is contained fully inside a critical radius $R \sim 2M/p_\mathrm{T}$. This radiation pattern differs significantly from QCD jets. This suggests that isolation criteria such as those used for hadronic τ-lepton decays may be used to isolate a sample of hadronic gauge boson decays. The same report [323] finds that a more traditional substructure analysis yields a separation between a boosted gauge boson sample and quark or gluon jet background for p_T up to several tens of TeV.

The reconstruction and isolation of hyper-boosted objects with p_T of 10-20 TeV poses important requirements on the detector design. This unexplored kinematic regime requires the development of new algorithms and techniques.

6.2 Lepton Colliders

A high-energy e^+e^- collider can provide precise measurements of the interactions of the Higgs boson, SM gauge bosons, the top quark and possible new states [300, 324, 325]. An accurate reconstruction of hadronic final states is a prerequisite for a precise measurements of Higgs boson and top quark couplings [326, 327, 328].

Excellent jet clustering is required to take full advantage of the potential of the machine and detectors. This is considerably more challenging than at the previous generation of lepton colliders, especially in complex multijet final states with multiple energy scales [329], where the contribution of jet clustering to the mass resolution may be sizeable or even dominant [328].

Detailed detector designs [330] have been developed for the linear collider projects. They achieve excellent single-particle reconstruction with highly granular calorimeters [331, 332], and particle-flow algorithms [333]. This approach yields a superb response to the jet energy and jet substructure observables [328].

The collection of a clean sample of energetic jets offers excellent opportunities to study jet formation in an environment that is more readily calculable, less affected by pileup, and much less prone to biases in trigger and selection. The detailed understanding of fragmentation would benefit analyses employing jet substructure at hadron colliders [334].

At linear colliders $\gamma\gamma \to$ hadrons production yields a diffuse background superposed on the signal [335]. Longitudinally invariant algorithms developed for hadron colliders are found to be significantly more robust than traditional e^+e^- jet reconstruction algorithms [303]. The same background resilience can be achieved with a classical e^+e^- inter-particle distance criterion by modifying the beam distance [327, 328].

Work in this direction has only just started [328] and only a few studies exist in the literature on the use of boosted object tagging and jet substructure analysis at future lepton colliders. The separation between quark and gluon jets that can be achieved with jet substructure is likely to play an important role in the isolation of a $H \to gg$ signal at a Higgs factory with $\sqrt{s} = 250$ GeV. The reconstruction of boosted gauge bosons, Higgs bosons and top quarks in e^+e^- colliders with a center-of-mass above 1 TeV can benefit from reconstruction, grooming and tagging algorithms developed for the LHC [328].

7 Conclusions

Jet substructure is a term used to describe the calculations, algorithms, and analysis techniques developed over the last decade and reviewed in this article. These methods are used to exploit the details of hadronic activity detectable by modern particle detectors such as ATLAS and CMS, and measurements and searches at both these experiments increasingly rely on one or more of the tools developed by the jet substructure community for boosted particle tagging and pileup suppression. Our hope is that this review serves as a useful reference, facilitating the community to continue tackling the challenges of high-energy experimental particle physics.

Acknowledgments

gesting this jet substructure review article. We also thank Andrew Larkoski and Ian Moult for the collaboration on the theoretical section of the review.

References

[1] LHCb Collaboration. Study of J/ψ Production in Jets. *Phys. Rev. Lett.*, 118:192001, May 2017. doi: 10.1103/PhysRevLett.118.192001. https://link.aps.org/doi/10.1103/PhysRevLett.118.192001.

[2] ALICE Collaboration. Jet-like correlations with neutral pion triggers in pp and central PbÃ¢ÂÅÂÅÂŞPb collisions at 2.76 TeV. *Phys. Lett. B*, 763:238 – 250, 2016. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2016.10.048. http://www.sciencedirect.com/science/article/pii/S0370269316306190.

[3] ALICE Collaboration. First measurement of jet mass in PbÃ¢ÂÅÂÅÂŞPb and pÃ¢ÂÅÂÅÂŞPb collisions at the LHC. *Phys. Lett. B*, 776:249 – 264, 2018. doi: https://doi.org/10.1016/j.physletb.2017.11.044. http://www.sciencedirect.com/science/article/pii/S037026931730936X.

[4] Boost 2011 Participants. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. https://arxiv.org/abs/1201.0008.

[5] Boost 2012 Participants. Boosted objects and jet substructure at the LHC. https://arxiv.org/abs/1311.2708.

[6] Boost 2013 Participants. Towards an Understanding of the Correlations in Jet Substructure. https://arxiv.org/abs/1504.00679.

[7] Andrew J. Larkoski, Ian Moult, and Benjamin Nachman. Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning. https://arxiv.org/abs/1709.04464.

[8] Gavin P. Salam. Towards Jetography. *Eur. Phys. J.*, C67:637–686, 2010. doi: 10.1140/epjc/s10052-010-1314-6. https://arxiv.org/abs/0906.1833.

[9] R. Keith Ellis, W. James Stirling, and B. R. Webber. QCD and collider physics. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 8:1–435, 1996.

[10] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003. http://inspirehep.net/record/796888.

[11] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004. http://inspirehep.net/record/796887.

[12] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9:P10009, 2014. doi: 10.1088/1748-0221/9/10/P10009. https://arxiv.org/abs/1405.6569.

[13] ATLAS Collaboration. ATLAS detector and physics performance: Technical Design Report, 1. 1999. https://cds.cern.ch/record/391176.

[14] ATLAS Collaboration. Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC. *Eur. Phys. J.*, C76(9):502, 2016. doi: 10.1140/epjc/s10052-016-4335-y. https://arxiv.org/abs/1606.01133.

[15] ATLAS Collaboration. Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul 2015. http://cds.cern.ch/record/2037683.

[16] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063. https://arxiv.org/abs/0802.1189.

[17] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017. doi: 10.1140/epjc/s10052-017-5004-5. https://arxiv.org/abs/1603.02934.

[18] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J.*, C77(7):466, 2017. doi: 10.1140/epjc/s10052-017-5031-2. https://arxiv.org/abs/1703.10485.

[19] ATLAS Collaboration. A neural network clustering algorithm for the ATLAS silicon pixel detector.

JINST, 9:P09009, 2014. doi: 10.1088/1748-0221/9/09/P09009. https://arxiv.org/abs/1406.7690.

[20] ATLAS Collaboration. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J.*, C77(10):673, 2017. doi: 10.1140/epjc/s10052-017-5225-7. https://arxiv.org/abs/1704.07983.

[21] ATLAS Collaboration. Measurement of track reconstruction inefficiencies in the core of jets via pixel dE/dx with the ATLAS experiment using $\sqrt{s} = 13$ TeV *pp* collision data. ATLAS PUB Note ATL-PHYS-PUB-2016-007, 2016. https://cds.cern.ch/record/2140460.

[22] ATLAS Collaboration. Modelling of Track Reconstruction Inside Jets with the 2016 ATLAS $\sqrt{s} = 13$ TeV *pp* Dataset. ATLAS PUB Note ATL-PHYS-PUB-2017-016, 2017. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-016/.

[23] ATLAS Collaboration. Measurement of performance of the pixel neural network clustering algorithm of the ATLAS experiment at $\sqrt{s} = 13$ TeV, 2015. https://cds.cern.ch/record/2054921.

[24] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. https://cds.cern.ch/record/1194487.

[25] CMS Collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002, 2010. https://cds.cern.ch/record/1279341.

[26] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003. https://arxiv.org/abs/1706.04965.

[27] CMS Collaboration. V Tagging Observables and Correlations. CMS Physics Analysis Summary CMS-PAS-JME-14-002, 2014. http://cds.cern.ch/record/1754913.

[28] CMS Collaboration. High p_{T} jet tracking with JetCore tracking and pixel cluster splitting. CMS DP , 2014. https://twiki.cern.ch/twiki/bin/view/CMSPublic/HighPtTrackingDP.

[29] ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J.*, C75:17, 2015. doi: 10.1140/epjc/s10052-014-3190-y. https://arxiv.org/abs/1406.0076.

[30] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D96(7):072002, 2017. doi: 10.1103/PhysRevD.96.072002. https://arxiv.org/abs/1703.09665.

[31] CMS Collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6:P11002, 2011. doi: 10.1088/1748-0221/6/11/P11002. https://arxiv.org/abs/1107.4277.

[32] Aviv Cukierman and Benjamin Nachman. Mathematical Properties of Numerical Inversion for Jet Calibrations. *Nucl. Instrum. Meth.*, A858:1–11, 2017. doi: 10.1016/j.nima.2017.03.038. https://arxiv.org/abs/1609.05195.

[33] ATLAS Collaboration. Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV. ATLAS CONF Note ATLAS-CONF-2015-002, 2015. https://cds.cern.ch/record/2001682.

[34] ATLAS Collaboration. A measurement of the calorimeter response to single hadrons and determination of the jet energy scale uncertainty using LHC Run-1 pp-collision data with the ATLAS detector. *Eur. Phys. J.*, C77, 2017. doi: 10.1140/epjc/s10052-016-4580-0. https://arxiv.org/abs/1607.08842.

[35] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: 10.1088/1748-0221/12/02/P02014. https://arxiv.org/abs/1607.03663.

[36] CMS Collaboration. Jet energy scale uncertainty correlations between ATLAS and CMS. CMS Physics Analysis Summary CMS-PAS-JME-14-003, 2014. https://cds.cern.ch/record/1967369.

[37] ATLAS Collaboration. Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76 (3):154, 2016. doi: 10.1140/epjc/s10052-016-3978-z. https://arxiv.org/abs/1510.05821.

[38] ATLAS Collaboration. Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS. ATLAS PUB Note ATL-PHYS-PUB-2015-033, 2015. https://cds.cern.ch/record/2041461.

[39] ATLAS Collaboration. Impact of Alternative Inputs and Grooming Methods on Large-R Jet Reconstruction in ATLAS. ATLAS PUB Note ATL-PHYS-PUB-2017-020, 2017. https://cds.cern.ch/record/2297485.

[40] ATLAS Collaboration. Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 06: 093, 2016. doi: 10.1007/JHEP06(2016)093. https://arxiv.org/abs/1603.03127.

[41] ATLAS Collaboration. Boosted hadronic top identification at ATLAS for early 13 TeV data. ATLAS PUB Note ATL-PHYS-PUB-2015-053, 2015. https://cds.cern.ch/record/2116351.

[42] ATLAS Collaboration. A measurement of the soft-drop jet mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2017. https://arxiv.org/abs/1711.08341.

[43] ATLAS Collaboration. Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *New J. Phys.*, 16(11):113013, 2014. doi: 10.1088/1367-2630/16/11/113013. https://arxiv.org/abs/1407.0800.

[44] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010. doi: 10.1007/JHEP02(2010)084. https://arxiv.org/abs/0912.1342.

[45] Benjamin Nachman, Pascal Nef, Ariel Schwartzman, Maximilian Swiatlowski, and Chaowaroj Wanotayaroj. Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC. *JHEP*,

02:075, 2015. doi: 10.1007/JHEP02(2015)075. https://arxiv.org/abs/1407.2922.

[46] ATLAS Collaboration. Search for pair production of gluinos decaying via stop and sbottom in events with b-jets and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D94(3):032003, 2016. doi: 10.1103/PhysRevD.94.032003. https://arxiv.org/abs/1605.09318.

[47] ATLAS Collaboration. Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Phys. Rev.*, D94 (5):052009, 2016. doi: 10.1103/PhysRevD.94.052009. https://arxiv.org/abs/1606.03903.

[48] ATLAS Collaboration. Search for new phenomena with large jet multiplicities and missing transverse momentum using large-radius jets and flavour-tagging at ATLAS in 13 TeV pp collisions. *JHEP*, 12: 034, 2017. doi: 10.1007/JHEP12(2017)034. https://arxiv.org/abs/1708.02794.

[49] ATLAS Collaboration. Jet reclustering and close-by effects in ATLAS Run 2. ATLAS CONF Note ATLAS-CONF-2017-062, 2017. http://cds.cern.ch/record/2275649.

[50] CMS Collaboration. Identification techniques for highly boosted W bosons that decay into hadrons. *JHEP*, 12:017, 2014. doi: 10.1007/JHEP12(2014)017. https://arxiv.org/abs/1410.4227.

[51] CMS Collaboration. Boosted Top Jet Tagging at CMS. CMS Physics Analysis Summary CMS-PAS-JME-13-007, 2014. http://cds.cern.ch/record/1647419.

[52] CMS Collaboration. Top Tagging with New Approaches. CMS Physics Analysis Summary CMS-PAS-JME-15-002, 2016. http://cds.cern.ch/record/2126325.

[53] CMS Collaboration. Studies of jet mass in dijet and W/Z + jet events. *JHEP*, 05:090, 2013. doi: 10.1007/JHEP05(2013)090. https://arxiv.org/abs/1303.4811.

[54] CMS Collaboration. Measurement of the differential jet production cross section with respect to jet mass and transverse momentum in dijet events from

pp collisions at $\sqrt{s}$ = 13 TeV. CMS Physics Analysis Summary CMS-PAS-SMP-16-010, 2017. http://cds.cern.ch/record/2273393.

[55] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010. doi: 10.1103/PhysRevD.81.094023. https://arxiv.org/abs/0912.0033.

[56] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013. doi: 10.1007/JHEP09(2013)029. https://arxiv.org/abs/1307.0007.

[57] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014. doi: 10.1007/JHEP10(2014)059. https://arxiv.org/abs/1407.6013.

[58] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014. doi: 10.1007/JHEP05(2014)146. https://arxiv.org/abs/1402.2657.

[59] CMS Collaboration. Jet algorithms performance in 13 TeV data. CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017. https://cds.cern.ch/record/2256875.

[60] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011. doi: 10.1007/JHEP03(2011)015. https://arxiv.org/abs/1011.2268.

[61] Jesse Thaler and Ken Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 02:093, 2012. doi: 10.1007/JHEP02(2012)093. https://arxiv.org/abs/1108.2701.

[62] ATLAS Collaboration. Jet mass reconstruction with the ATLAS Detector in early Run 2 data. ATLAS CONF Note ATLAS-CONF-2016-035, 2016. http://cds.cern.ch/record/2200211.

[63] ATLAS Collaboration. Measurement of large radius jet mass reconstruction performance at $\sqrt{s}$ = 8 tev using the atlas detector. ATLAS CONF Note ATLAS-CONF-2016-008, 2016. https://cds.cern.ch/record/2139642.

[64] ATLAS Collaboration. In-situ measurements of large-radius jet reconstruction performance. ATLAS CONF Note ATLAS-CONF-2017-063, 2017. http://cds.cern.ch/record/2275655.

[65] ATLAS Collaboration. Improving jet substructure performance in atlas using track-caloclusters. ATLAS PUB Note ATL-PHYS-PUB-2017-015, 2017. https://cds.cern.ch/record/2275636.

[66] CMS Collaboration. Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at sqrt(s) = 13 TeV. 2017. https://arxiv.org/abs/1708.05379.

[67] CMS Collaboration. Measurement of the jet mass in highly boosted t$\bar{\text{t}}$ events from pp collisions at $\sqrt{s}$ = 8 TeV. *Eur. Phys. J.*, C77(7):467, 2017. doi: 10.1140/epjc/s10052-017-5030-3. https://arxiv.org/abs/1703.06330.

[68] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The catchment area of jets. *JHEP*, 04:005, 2008. doi: 10.1088/1126-6708/2008/04/005. https://arxiv.org/abs/0802.1188.

[69] CMS Collaboration. Pileup Removal Algorithms. CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014. http://cds.cern.ch/record/1751454.

[70] ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in *pp* collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector. *Eur. Phys. J.*, C76(11):581, 2016. doi: 10.1140/epjc/s10052-016-4395-z. https://arxiv.org/abs/1510.03823.

[71] Gregory Soyez, Gavin P. Salam, Jihun Kim, Souvik Dutta, and Matteo Cacciari. Pileup subtraction for jet shapes. *Phys. Rev. Lett.*, 110(16):162001, 2013. doi: 10.1103/PhysRevLett.110.162001. https://arxiv.org/abs/1211.2811.

[72] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. SoftKiller, a particle-level pileup removal method. *Eur. Phys. J.*, C75(2):59, 2015. doi: 10.1140/epjc/s10052-015-3267-2. https://arxiv.org/abs/1407.0408.

[73] Peter Berta, Martin Spousta, David W. Miller, and Rupert Leitner. Particle-level pileup subtraction for

jets and jet shapes. *JHEP*, 06:092, 2014. doi: 10.1007/JHEP06(2014)092. https://arxiv.org/abs/1403.3108.

[74] CMS Collaboration. Pileup Jet Identification. CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013. https://cds.cern.ch/record/1581583.

[75] ATLAS Collaboration. Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector. *Eur. Phys. J.*, C77(9):580, 2017. doi: 10.1140/epjc/s10052-017-5081-5. https://arxiv.org/abs/1705.02211.

[76] ATLAS Collaboration. Event displays from run 2 physics analyses, 2016. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Physics.

[77] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz. Pileup Mitigation with Machine Learning (PUMML). *JHEP*, 12:051, 2017. doi: 10.1007/JHEP12(2017)051. https://arxiv.org/abs/1707.08600.

[78] Daniele Bertolini, Tucker Chan, and Jesse Thaler. Jet Observables Without Jet Algorithms. *JHEP*, 04:013, 2014. doi: 10.1007/JHEP04(2014)013. https://arxiv.org/abs/1310.7584.

[79] David Krohn, Matthew D. Schwartz, Matthew Low, and Lian-Tao Wang. Jet Cleansing: Pileup Removal at High Luminosity. *Phys. Rev.*, D90(6):065020, 2014. doi: 10.1103/PhysRevD.90.065020. https://arxiv.org/abs/1309.4777.

[80] D. Contardo, M. Klute, J. Mans, L. Silvestris, and J. Butler. Technical Proposal for the Phase-II Upgrade of the CMS Detector, 2015. https://cds.cern.ch/record/2020886.

[81] ATLAS Collaboration. Constituent-level pileup mitigation performance using 2015 data. ATLAS CONF Note ATLAS-CONF-2017-065, 2017. https://cds.cern.ch/record/2281055.

[82] W. Bartel et al., JADE Collaboration. Observation of planar three jet events in e^+e^- annihilation and evidence for gluon bremsstrahlung. *Phys. Lett.*, B91:142, 1980. doi: 10.1016/0370-2693(80)90680-2. http://inspirehep.net/record/143985.

[83] Ch. Berger et al., PLUTO Collaboration. Evidence for Gluon Bremsstrahlung in e^+e^- Annihilations at High-Energies. *Phys. Lett.*, B86:418, 1979. doi: 10.1016/0370-2693(79)90869-4. http://inspirehep.net/record/142632.

[84] D.P Barber et al., MARK-J Collaboration. Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies. *Phys. Rev. Lett.*, 43:830, 1979. doi: 10.1103/PhysRevLett.43.830. http://inspirehep.net/record/142085.

[85] R. Brandelik et al., TASSO Collaboration. Evidence for Planar Events in e^+e^- Annihilation at High-Energies. *Phys. Lett.*, B86:243, 1979. doi: 10.1016/0370-2693(79)90830-X. http://inspirehep.net/record/142123.

[86] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298–318, 1977. doi: 10.1016/0550-3213(77)90384-4. http://inspirehep.net/record/119585.

[87] Andrea Banfi, Gavin P. Salam, and Giulia Zanderighi. Infrared safe definition of jet flavor. *Eur. Phys. J.*, C47:113–124, 2006. doi: 10.1140/epjc/s2006-02552-4. https://arxiv.org/abs/hep-ph/0601139.

[88] Andy Buckley and Chris Pollard. QCD-aware partonic jet clustering for truth-jet flavour labelling. *Eur. Phys. J.*, C76(2):71, 2016. doi: 10.1140/epjc/s10052-016-3925-z. https://arxiv.org/abs/1507.00508.

[89] Christopher Frye, Andrew J. Larkoski, Matthew D. Schwartz, and Kai Yan. Factorization for groomed jet substructure beyond the next-to-leading logarithm. *JHEP*, 07:064, 2016. doi: 10.1007/JHEP07(2016)064. https://arxiv.org/abs/1603.09338.

[90] Jason Gallicchio and Matthew D. Schwartz. Quark and Gluon Tagging at the LHC. *Phys. Rev. Lett.*, 107:172001, 2011. doi: 10.1103/PhysRevLett.107.172001. https://arxiv.org/abs/1106.3076.

[91] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy Correlation Functions for Jet Substructure. *JHEP*, 06:108, 2013. doi: 10.1007/JHEP06(2013)108. https://arxiv.org/abs/1305.0007.

[92] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. Gaining (Mutual) Information about Quark/Gluon Discrimination. *JHEP*, 11:129, 2014.

doi: 10.1007/JHEP11(2014)129. https://arxiv.org/abs/1408.3122.

[93] Christopher Frye, Andrew J. Larkoski, Jesse Thaler, and Kevin Zhou. Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables. *JHEP*, 09:083, 2017. doi: 10.1007/JHEP09(2017)083. https://arxiv.org/abs/1704.06266.

[94] Patrick T. Komiske, Eric M. Metodiev, and Matthew D. Schwartz. Deep learning in color: towards automated quark/gluon jet discrimination. *JHEP*, 01:110, 2016. doi: 10.1007/JHEP01(2017)110. https://arxiv.org/abs/1612.01551.

[95] Philippe Gras, Stefan Höche, Deepak Kar, Andrew Larkoski, Leif Lönnblad, Simon Plätzer, Andrzej Siódmok, Peter Skands, Gregory Soyez, and Jesse Thaler. Systematics of quark/gluon tagging. *JHEP*, 07:091, 2017. doi: 10.1007/JHEP07(2017)091. https://arxiv.org/abs/1704.03878.

[96] ATLAS Collaboration. Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J.*, C74:3023, 2014. doi: 10.1140/epjc/s10052-014-3023-z. http://arxiv.org/abs/arXiv:1405.6583.

[97] ATLAS Collaboration. Discrimination of Light Quark and Gluon Jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. ATLAS CONF Note ATLAS-CONF-2016-034, 2016. http://cds.cern.ch/record/2200202.

[98] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector. ATLAS PUB Note ATL-PHYS-PUB-2017-009, 2017. https://cds.cern.ch/record/2263679.

[99] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector. ATLAS PUB Note ATL-PHYS-PUB-2017-017, 2017. http://cds.cern.ch/record/2275641.

[100] CMS Collaboration. Performance of quark/gluon discrimination in 8 TeV pp data. CMS Physics Analysis Summary CMS-PAS-JME-13-002, 2013. https://cds.cern.ch/record/1599732.

[101] CMS Collaboration. Performance of quark/gluon discrimination in 13 TeV data. CMS Detector Performance Summary CMS-DP-2016-070, 2016. https://cds.cern.ch/record/2234117.

[102] CMS Collaboration. New Developments for Jet Substructure Reconstruction in CMS. Cms detector performance summary, 2017. https://cds.cern.ch/record/2275226.

[103] CMS Collaboration. Search for a Higgs boson in the decay channel $H \to ZZ^{(*)} \to qqbar\ell^-l^+$ in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1204:036, 2012. doi: 10.1007/JHEP04(2012)036. https://arxiv.org/abs/1202.1416.

[104] CMS Collaboration. Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 10:062, 2013. doi: 10.1007/JHEP10(2013)062. https://arxiv.org/abs/1305.7389.

[105] CMS Collaboration. Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$. *Phys. Rev.*, D92(3):032008, 2015. doi: 10.1103/PhysRevD.92.032008. https://arxiv.org/abs/1506.01010.

[106] CMS Collaboration. Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 10:005, 2017. doi: 10.1007/JHEP10(2017)005. https://arxiv.org/abs/1707.03316.

[107] ATLAS Collaboration. Search for the Standard Model Higgs boson produced by vector-boson fusion and decaying to bottom quarks in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector. *JHEP*, 11:112, 2016. doi: 10.1007/JHEP11(2016)112. https://arxiv.org/abs/1606.02181.

[108] ATLAS Collaboration. Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 12:055, 2015. doi: 10.1007/JHEP12(2015)055. https://arxiv.org/abs/1506.00962.

[109] Lucio Mwinmaarong Dery, Benjamin Nachman, Francesco Rubbo, and Ariel Schwartzman. Weakly Supervised Classification in High Energy Physics. *JHEP*,

05:145, 2017. doi: 10.1007/JHEP05(2017)145. https://arxiv.org/abs/1702.00414.

[110] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. Classification without labels: Learning from mixed samples in high energy physics. *JHEP*, 10: 174, 2017. doi: 10.1007/JHEP10(2017)174. https://arxiv.org/abs/1708.02949.

[111] CMS Collaboration. Study of Pileup Removal Algorithms for Jets. 2014. https://cds.cern.ch/record/1751454.

[112] Stephen D. Ellis, Andrew Hornig, Tuhin S. Roy, David Krohn, and Matthew D. Schwartz. Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure. *Phys. Rev. Lett.*, 108:182003, 2012. doi: 10.1103/PhysRevLett.108.182003. https://arxiv.org/abs/1201.1914.

[113] Andrew J. Larkoski, Ian Moult, and Duff Neill. Power Counting to Better Jet Observables. *JHEP*, 12: 009, 2014. doi: 10.1007/JHEP12(2014)009. https://arxiv.org/abs/1409.6298.

[114] Andrew J. Larkoski, Ian Moult, and Duff Neill. Analytic Boosted Boson Discrimination. *JHEP*, 05: 117, 2016. doi: 10.1007/JHEP05(2016)117. https://arxiv.org/abs/1507.03018.

[115] Ian Moult, Lina Necib, and Jesse Thaler. New Angles on Energy Correlation Functions. *JHEP*, 12: 153, 2016. doi: 10.1007/JHEP12(2016)153. https://arxiv.org/abs/1609.07483.

[116] Leandro G. Almeida, Seung J. Lee, Gilad Perez, George F. Sterman, Ilmo Sung, and Joseph Virzi. Substructure of high-p_T Jets at the LHC. *Phys. Rev.*, D79:074017, 2009. doi: 10.1103/PhysRevD.79.074017. https://arxiv.org/abs/0807.0234.

[117] J. M. Butterworth, B. E. Cox, and Jeffrey R. Forshaw. WW scattering at the CERN LHC. *Phys. Rev.*, D65:096014, 2002. doi: 10.1103/PhysRevD.65.096014. https://arxiv.org/abs/hep-ph/0201098.

[118] Jesse Thaler and Lian-Tao Wang. Strategies to Identify Boosted Tops. *JHEP*, 07:092, 2008. doi: 10.1088/1126-6708/2008/07/092. https://arxiv.org/abs/0806.0023.

[119] Jason Gallicchio and Matthew D. Schwartz. Seeing in Color: Jet Superstructure. *Phys. Rev. Lett.*, 105: 022001, 2010. doi: 10.1103/PhysRevLett.105.022001. https://arxiv.org/abs/1001.5027.

[120] ATLAS Collaboration. Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV. ATLAS PUB Note ATL-PHYS-PUB-2017-004, 2017. https://cds.cern.ch/record/2259646.

[121] Andrew J. Larkoski, Duff Neill, and Jesse Thaler. Jet Shapes with the Broadening Axis. *JHEP*, 04: 017, 2014. doi: 10.1007/JHEP04(2014)017. https://arxiv.org/abs/1401.2158.

[122] CMS Collaboration. Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 01:097, 2018. doi: 10.1007/JHEP01(2018)097. https://arxiv.org/abs/1710.00159.

[123] CMS Collaboration. Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV. *JHEP*, 08:174, 2014. doi: 10.1007/JHEP08(2014)174. https://arxiv.org/abs/1405.3447.

[124] CMS Collaboration. Search for massive WH resonances decaying into the $\ell\nu b\bar{b}$ final state at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76(5):237, 2016. doi: 10.1140/epjc/s10052-016-4067-z. https://arxiv.org/abs/1601.06431.

[125] CMS Collaboration. Search for massive resonances decaying into WW, WZ or ZZ bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:162, 2017. doi: 10.1007/JHEP03(2017)162. https://arxiv.org/abs/1612.09159.

[126] ATLAS Collaboration. Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B774:494–515, 2017. doi: 10.1016/j.physletb.2017.09.066. https://arxiv.org/abs/1707.06958.

[127] ATLAS Collaboration. Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B777:

91–113, 2018. doi: 10.1016/j.physletb.2017.12.011. https://arxiv.org/abs/1708.04445.

[128] ATLAS Collaboration. Search for WW/WZ resonance production in $\ell\nu qq$ final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. ATLAS CONF Note ATLAS-CONF-2017-051, 2017. http://cds.cern.ch/record/2273867.

[129] James Dolen, Philip Harris, Simone Marzani, Salvatore Rappoccio, and Nhan Tran. Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure. *JHEP*, 05:156, 2016. doi: 10.1007/JHEP05(2016)156. https://arxiv.org/abs/1603.00027.

[130] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Sogaard. Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. *Phys. Rev.*, D96 (7):074034, 2017. doi: 10.1103/PhysRevD.96.074034. https://arxiv.org/abs/1703.03507.

[131] J. A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra. A generic anti-QCD jet tagger. *JHEP*, 11:163, 2017. doi: 10.1007/JHEP11(2017)163. https://arxiv.org/abs/1709.01087.

[132] Ian Moult, Benjamin Nachman, and Duff Neill. Convolved Substructure: Analytically Decorrelating Jet Substructure Observables. 2017. https://arxiv.org/abs/1710.06859.

[133] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with Variable R. *JHEP*, 06:059, 2009. doi: 10.1088/1126-6708/2009/06/059. https://arxiv.org/abs/0903.0392.

[134] ATLAS Collaboration. Boosted Object Tagging with Variable-R Jets in the ATLAS Detector. ATLAS PUB Note ATL-PHYS-PUB-2016-013, 2016. https://cds.cern.ch/record/2199360.

[135] ATLAS Collaboration. A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector. *Eur. Phys. J.*, C76 (5):238, 2016. doi: 10.1140/epjc/s10052-016-4065-1. https://arxiv.org/abs/1509.04939.

[136] ATLAS Collaboration. Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector.

ATLAS PUB Note ATL-PHYS-PUB-2014-013, 2014. https://cds.cern.ch/record/1750681.

[137] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks. *Phys. Rev. Lett.*, 101:142001, 2008. doi: 10.1103/PhysRevLett.101.142001. https://arxiv.org/abs/0806.0848.

[138] CMS Collaboration. A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. CMS Physics Analysis Summary CMS-PAS-JME-09-001, 2009. http://cds.cern.ch/record/1194489.

[139] CMS Collaboration. Jet Substructure Algorithms. CMS Physics Analysis Summary CMS-PAS-JME-10-013, 2011. http://cds.cern.ch/record/1333700.

[140] Tilman Plehn, Gavin P. Salam, and Michael Spannowsky. Fat Jets for a Light Higgs. *Phys. Rev. Lett.*, 104:111801, 2010. doi: 10.1103/PhysRevLett.104.111801. https://arxiv.org/abs/0910.5472.

[141] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. Stop Reconstruction with Tagged Tops. *JHEP*, 10:078, 2010. doi: 10.1007/JHEP10(2010)078. https://arxiv.org/abs/1006.2833.

[142] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 09:076, 2013. doi: 10.1007/JHEP09(2013)076. https://arxiv.org/abs/1306.4945.

[143] Davison E. Soper and Michael Spannowsky. Finding physics signals with shower deconstruction. *Phys. Rev.*, D84:074002, 2011. doi: 10.1103/PhysRevD.84.074002. https://arxiv.org/abs/1102.3480.

[144] Davison E. Soper and Michael Spannowsky. Finding top quarks with shower deconstruction. *Phys. Rev.*, D87:054012, 2013. doi: 10.1103/PhysRevD.87.054012. https://arxiv.org/abs/1211.3140.

[145] ATLAS Collaboration. Search for $W' \to tb$ decays in the hadronic final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018. https://arxiv.org/abs/1801.07893.

[146] ATLAS Collaboration. Performance of top quark and w boson tagging in run 2 with atlas. *ATLAS-CONF-2017-064*, 2017. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ ATLAS-CONF-2017-064/.

[147] Gregor Kasieczka, Tilman Plehn, Torben Schell, Thomas Strebler, and Gavin P. Salam. Resonance Searches with an Updated Top Tagger. *JHEP*, 06:203, 2015. doi: 10.1007/JHEP06(2015)203. https://arxiv.org/abs/1503.05921.

[148] ATLAS Collaboration. Top and Boson Tagger Background Performance - Moriond2017. Technical Report ATL-JETM-2017-005, 2017. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/ JETM-2017-005/.

[149] Tobias Lapsien, Roman Kogler, and Johannes Haller. A new tagger for hadronically decaying heavy particles at the LHC. *Eur. Phys. J.*, C76(11):600, 2016. doi: 10.1140/epjc/s10052-016-4443-8. https://arxiv.org/abs/1606.04961.

[150] 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*, 08:148, 2015. doi: 10.1007/JHEP08(2015)148. https://arxiv.org/abs/1505.07018.

[151] ATLAS Collaboration. Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector. 2017. https://arxiv.org/abs/1711.11520.

[152] CMS Collaboration. Search for supersymmetry using hadronic top quark tagging in 13 TeV pp collisions. CMS Physics Analysis Summary CMS-PAS-SUS-16-050, 2017. http://cds.cern.ch/record/2262651.

[153] CMS Collaboration. W and top tagging scale factors. CMS Detector Performance Summary CMS-DP-2017-026, 2017. https://cds.cern.ch/record/2275225.

[154] 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, 2016. http://cds.cern.ch/record/2138504.

[155] ATLAS Collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. ATLAS PUB Note ATL-PHYS-PUB-2016-012, 2016. https://cds.cern.ch/record/2160731.

[156] CMS Collaboration. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013. https://arxiv.org/abs/1211.4462.

[157] CMS Collaboration. Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013. http://cds.cern.ch/record/1581306.

[158] ATLAS Collaboration. Studies of b-tagging performance and jet substructure in a high p_T $g \to b\bar{b}$ rich sample of large-R jets from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. ATLAS CONF Note ATLAS-CONF-2016-002, 2016. http://cds.cern.ch/record/2135187.

[159] ATLAS Collaboration. Performance of b-Jet Identification in the ATLAS Experiment. *JINST*, 11(04):P04008, 2016. doi: 10.1088/1748-0221/11/04/P04008. https://arxiv.org/abs/1512.01094.

[160] ATLAS Collaboration. Boosted Higgs ($\to b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV. ATLAS CONF Note ATLAS-CONF-2016-039, 2016. http://cds.cern.ch/record/2206038.

[161] ATLAS Collaboration. b-tagging in dense environments. ATLAS PUB Note ATL-PHYS-PUB-2014-014, 2014. https://cds.cern.ch/record/1750682.

[162] CMS Collaboration. Identification of double-b quark jets in boosted event topologies. CMS Physics Analysis Summary CMS-PAS-BTV-15-002, 2016. http://cds.cern.ch/record/2195743.

[163] CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. Submitted to *JINST*, 2017. https://arxiv.org/abs/1712.07158.

[164] ATLAS Collaboration. Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\to b\bar{b}$) Tagging in ATLAS. ATLAS PUB Note ATL-PHYS-PUB-2017-010, 2017. https://cds.cern.ch/record/2268678.

[165] ATLAS Collaboration. Identification and Tagging of Double b-hadron jets with the ATLAS Detector. ATLAS CONF Note ATLAS-CONF-2012-100, 2012. http://cds.cern.ch/record/1462603.

[166] CMS Collaboration. Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s}$ = 7 TeV. *Eur. Phys. J.*, C73(3):2339, 2013. doi: 10.1140/epjc/s10052-013-2339-4. https://arxiv.org/abs/1211.2220.

[167] CMS Collaboration. Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s}$ = 8 TeV. *Eur. Phys. J.*, C75(11):542, 2015. doi: 10.1140/epjc/s10052-015-3709-x. https://arxiv.org/abs/1505.04480.

[168] CMS Collaboration. Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV. *Phys. Rev.*, D95(9):092001, 2017. doi: 10.1103/PhysRevD.95.092001. https://arxiv.org/abs/1610.04191.

[169] ATLAS Collaboration. Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s}$ = 7 TeV in the ATLAS detector. *JHEP*, 06:100, 2015. doi: 10.1007/JHEP06(2015)100. https://arxiv.org/abs/1502.05923.

[170] ATLAS Collaboration. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector. *Eur. Phys. J.*, C76(10):538, 2016. doi: 10.1140/epjc/s10052-016-4366-4. https://arxiv.org/abs/1511.04716.

[171] ATLAS Collaboration. Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector. *Eur. Phys. J.*, C77(5):292, 2017. doi: 10.1140/epjc/s10052-017-4821-x. https://arxiv.org/abs/1612.05220.

[172] ATLAS Collaboration. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector. *JHEP*, 11:191, 2017. doi: 10.1007/JHEP11(2017)191. https://arxiv.org/abs/1708.00727.

[173] ATLAS Collaboration. Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector. *Phys. Rev.*, D93(3):032009, 2016. doi: 10.1103/PhysRevD.93.032009. https://arxiv.org/abs/1510.03818.

[174] CMS Collaboration. Measurement of the integrated and differential $t\bar{t}$ production cross sections for high-p_t top quarks in pp collisions at $\sqrt{s}$ = 8 TeV. *Phys. Rev.*, D94(7):072002, 2016. doi: 10.1103/PhysRevD.94.072002. https://arxiv.org/abs/1605.00116.

[175] CMS Collaboration. Measurement of the differential $t\bar{t}$ cross section with high-p_T top-quark jets in the all-hadronic channel at $\sqrt{s}$ = 8 TeV. CMS Physics Analysis Summary CMS-PAS-TOP-16-018, 2017. https://cds.cern.ch/record/2280473.

[176] ATLAS Collaboration. Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector. 2018. https://arxiv.org/abs/1801.02052.

[177] S. Brandt, C. Peyrou, R. Sosnowski, and A. Wroblewski. The Principal axis of jets. An Attempt to analyze high-energy collisions as two-body processes. *Phys. Lett.*, 12:57–61, 1964. doi: 10.1016/0031-9163(64)91176-X.

[178] Edward Farhi. Quantum chromodynamics test for jets. *Phys. Rev. Lett.*, 39:1587–1588, Dec 1977. doi: 10.1103/PhysRevLett.39.1587. https://link.aps.org/doi/10.1103/PhysRevLett.39.1587.

[179] J. D. Bjorken and Stanley J. Brodsky. Statistical Model for electron-Positron Annihilation Into Hadrons. *Phys. Rev.*, D1:1416–1420, 1970. doi: 10.1103/PhysRevD.1.1416. https://inspirehep.net/record/54788.

[180] ATLAS Collaboration. Measurement of the cross section of high transverse momentum $Z \rightarrow b\bar{b}$ production in proton–proton collisions at $\sqrt{s}$ = $8 TeV$ with the ATLAS Detector. *Phys. Lett.*, B738:25–43, 2014. doi: 10.1016/j.physletb.2014.09.020. https://arxiv.org/abs/1404.7042.

[181] ATLAS Collaboration. Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector. *JHEP*, 12:024,

2017. doi: 10.1007/JHEP12(2017)024. https://arxiv.org/abs/1708.03299.

[182] CMS Collaboration. Evidence for the Higgs boson decay to a bottom quark-antiquark pair. 2017. https://arxiv.org/abs/1709.07497.

[183] CMS Collaboration. Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair. *Phys. Rev. Lett.*, 120(7):071802, 2018. doi: 10.1103/PhysRevLett.120.071802. https://arxiv.org/abs/1709.05543.

[184] ATLAS Collaboration. Study of Jet Shapes in Inclusive Jet Production in *pp* Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector. *Phys. Rev.*, D83:052003, 2011. doi: 10.1103/PhysRevD.83.052003. https://arxiv.org/abs/1101.0070.

[185] 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.*, C72:2211, 2012. doi: 10.1140/epjc/s10052-012-2211-y. https://arxiv.org/abs/1206.2135.

[186] ATLAS Collaboration. Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Eur. Phys. J.*, C73(12):2676, 2013. doi: 10.1140/epjc/s10052-013-2676-3. https://arxiv.org/abs/1307.5749.

[187] CMS Collaboration. Shape, Transverse Size, and Charged Hadron Multiplicity of Jets in pp Collisions at 7 TeV. *JHEP*, 06:160, 2012. doi: 10.1007/JHEP06(2012)160. https://arxiv.org/abs/1204.3170.

[188] ATLAS Collaboration. Properties of jets measured from tracks in proton-proton collisions at center-of-mass energy $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev.*, D84:054001, 2011. doi: 10.1103/PhysRevD.84.054001. https://arxiv.org/abs/1107.3311.

[189] ATLAS Collaboration. Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV *pp* collisions with the ATLAS detector. *Eur. Phys. J.*, C76(6):322, 2016. doi: 10.1140/epjc/s10052-016-4126-5. https://arxiv.org/abs/1602.00988.

[190] ATLAS Collaboration. Measurement of the jet fragmentation function and transverse profile in proton-proton collisions at a center-of-mass energy of 7 TeV with the ATLAS detector. *Eur. Phys. J.*, C71:1795, 2011. doi: 10.1140/epjc/s10052-011-1795-y. https://arxiv.org/abs/1109.5816.

[191] ATLAS Collaboration. Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV. *Phys. Lett.*, B750:475–493, 2015. doi: 10.1016/j.physletb.2015.09.051. https://arxiv.org/abs/1506.05629.

[192] ATLAS Collaboration. Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV *pp* collisions with the ATLAS experiment. *JHEP*, 05:128, 2012. doi: 10.1007/JHEP05(2012)128. https://arxiv.org/abs/1203.4606.

[193] ATLAS Collaboration. Measurement of kT splitting scales in W->lv events at sqrt(s)=7 TeV with the ATLAS detector. *Eur. Phys. J.*, C73(5):2432, 2013. doi: 10.1140/epjc/s10052-013-2432-8. https://arxiv.org/abs/1302.1415.

[194] ATLAS Collaboration. Measurement of jet charge in dijet events from $\sqrt{s}$=8 TeV pp collisions with the ATLAS detector. *Phys. Rev.*, D93(5):052003, 2016. doi: 10.1103/PhysRevD.93.052003. https://arxiv.org/abs/1509.05190.

[195] CMS Collaboration. Measurements of jet charge with dijet events in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 10:131, 2017. doi: 10.1007/JHEP10(2017)131. https://arxiv.org/abs/1706.05868.

[196] CMS Collaboration. Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV. *Phys. Rev.*, D93(7):072004, 2016. doi: 10.1103/PhysRevD.93.072004. https://arxiv.org/abs/1509.04044.

[197] ATLAS Collaboration. Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets and $t\bar{t} \to$ dilepton channels using $\sqrt{s} = 7$ TeV ATLAS data. *Eur. Phys. J.*, C75(7):330, 2015. doi: 10.1140/epjc/s10052-015-3544-0. https://arxiv.org/abs/1503.05427.

[198] ATLAS Collaboration. Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8$

TeV ATLAS data. *Phys. Lett.*, B761:350–371, 2016. doi: 10.1016/j.physletb.2016.08.042. https://arxiv.org/abs/1606.02179.

[199] Tevatron Electroweak Working Group. Combination of CDF and D0 results on the mass of the top quark using up to 9.7 fb^{-1} at the Tevatron. 2014. https://arxiv.org/abs/1407.2682.

[200] Andre H. Hoang, Sonny Mantry, Aditya Pathak, and Iain W. Stewart. Extracting a Short Distance Top Mass with Light Grooming. 2017. https://arxiv.org/abs/1708.02586.

[201] Anders Andreassen and Matthew D. Schwartz. Reducing the Top Quark Mass Uncertainty with Jet Grooming. *JHEP*, 10:151, 2017. doi: 10.1007/JHEP10(2017)151. https://arxiv.org/abs/1705.07135.

[202] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036. https://arxiv.org/abs/0710.3820.

[203] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev.*, D82:074018, 2010. doi: 10.1103/PhysRevD.82.074018. https://arxiv.org/abs/1005.3457.

[204] Wouter J. Waalewijn. Calculating the Charge of a Jet. *Phys. Rev.*, D86:094030, 2012. doi: 10.1103/PhysRevD.86.094030. https://arxiv.org/abs/1209.3019.

[205] David Krohn, Matthew D. Schwartz, Tongyan Lin, and Wouter J. Waalewijn. Jet Charge at the LHC. *Phys. Rev. Lett.*, 110(21):212001, 2013. doi: 10.1103/PhysRevLett.110.212001. https://arxiv.org/abs/1209.2421.

[206] CMS Collaboration. Search for Anomalous $t\bar{t}$ Production in the Highly-Boosted All-Hadronic Final State. *JHEP*, 09:029, 2012. doi: 10.1007/JHEP09(2012)029, 10.1007/JHEP03(2014)132. https://arxiv.org/abs/1204.2488. [Erratum: JHEP03,132(2014)].

[207] CMS Collaboration. Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys.*

Rev., D93(1):012001, 2016. doi: 10.1103/PhysRevD.93.012001. https://arxiv.org/abs/1506.03062.

[208] ATLAS Collaboration. Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of *pp* collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev.*, D88(1):012004, 2013. doi: 10.1103/PhysRevD.88.012004. https://arxiv.org/abs/1305.2756.

[209] 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*, 01:116, 2013. doi: 10.1007/JHEP01(2013)116. https://arxiv.org/abs/1211.2202.

[210] ATLAS Collaboration. Search for heavy particles decaying to pairs of highly-boosted top quarks using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. ATLAS CONF Note ATLAS-CONF-2016-014, 2016. http://cds.cern.ch/record/2141001.

[211] CMS Collaboration. Search for $t\bar{t}$ resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:001, 2017. doi: 10.1007/JHEP07(2017)001. https://arxiv.org/abs/1704.03366.

[212] CMS Collaboration. Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV. *Phys. Lett.*, B723:280–301, 2013. doi: 10.1016/j.physletb.2013.05.040. https://arxiv.org/abs/1212.1910.

[213] CMS Collaboration. Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 08:173, 2014. doi: 10.1007/JHEP08(2014)173. https://arxiv.org/abs/1405.1994.

[214] CMS Collaboration. Search for Narrow High-Mass Resonances in ProtonâĂŞProton Collisions at $\sqrt{s} = 8$ TeV Decaying to a Z and a Higgs Boson. *Phys. Lett.*, B748:255–277, 2015. doi: 10.1016/j.physletb.2015.07.011. https://arxiv.org/abs/1502.04994.

[215] CMS Collaboration. Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 02:145, 2016. doi: 10.1007/JHEP02(2016)145. https://arxiv.org/abs/1506.01443.

[216] CMS Collaboration. Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks. *Eur. Phys. J.*, C76(7):371, 2016. doi: 10.1140/epjc/s10052-016-4206-6. https://arxiv.org/abs/1602.08762.

[217] ATLAS Collaboration. Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector. *Phys. Lett.*, B755: 285–305, 2016. doi: 10.1016/j.physletb.2016.02.015. https://arxiv.org/abs/1512.05099.

[218] ATLAS Collaboration. Search for WW/WZ resonance production in $\ell\nu qq$ final states in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector. 2017. https://arxiv.org/abs/1710.07235.

[219] ATLAS Collaboration. Searches for heavy ZZ and ZW resonances in the $\ell\ell qq$ and $\nu\nu qq$ final states in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector. 2017. https://arxiv.org/abs/1708.09638.

[220] ATLAS Collaboration. Searches for heavy diboson resonances in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector. *JHEP*, 09:173, 2016. doi: 10.1007/JHEP09(2016)173. https://arxiv.org/abs/1606.04833.

[221] ATLAS Collaboration. Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} =$ 13 TeV pp collisions with the ATLAS detector. 2017. https://arxiv.org/abs/1712.06518.

[222] ATLAS Collaboration. Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} =$ 8 TeVwith the ATLAS detector. *Eur. Phys. J.*, C75(9):412, 2015. doi: 10.1140/epjc/s10052-015-3628-x. https://arxiv.org/abs/1506.00285.

[223] ATLAS Collaboration. Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector. *Phys. Rev.*, D94(5):052002, 2016. doi: 10.1103/PhysRevD.94.052002. https://arxiv.org/abs/1606.04782.

[224] CMS Collaboration. Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks. *Phys. Lett.*, B768:137–162, 2017. doi:

10.1016/j.physletb.2017.02.040. https://arxiv.org/abs/1610.08066.

[225] CMS Collaboration. Combination of searches for heavy resonances decaying to WW, WZ, ZZ, WH, and ZH boson pairs in proton-proton collisions at $\sqrt{s} =$ 8 and 13 TeV. *Phys. Lett.*, B774:533–558, 2017. doi: 10.1016/j.physletb.2017.09.083. https://arxiv.org/abs/1705.09171.

[226] CMS Collaboration. Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} =$ 13 TeV. *Eur. Phys. J.*, C77(9):636, 2017. doi: 10.1140/epjc/s10052-017-5192-z. https://arxiv.org/abs/1707.01303.

[227] CMS Collaboration. Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} =$ 13 TeV. CMS Physics Analysis Summary CMS-PAS-B2G-16-026, 2017. http://cds.cern.ch/record/2264684.

[228] CMS Collaboration. Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s}$ = 8 TeV. *Phys. Lett.*, B729:149–171, 2014. doi: 10.1016/j.physletb.2014.01.006. https://arxiv.org/abs/1311.7667.

[229] CMS Collaboration. Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure. *JHEP*, 06: 080, 2015. doi: 10.1007/JHEP06(2015)080. https://arxiv.org/abs/1503.01952.

[230] CMS Collaboration. Search for vector-like charge 2/3 T quarks in proton-proton collisions at sqrt(s) = 8 TeV. *Phys. Rev.*, D93(1):012003, 2016. doi: 10.1103/PhysRevD.93.012003. https://arxiv.org/abs/1509.04177.

[231] CMS Collaboration. Search for pair-produced vectorlike B quarks in proton-proton collisions at $\sqrt{s}$=8 âĂĽâĂĽTeV. *Phys. Rev.*, D93(11):112009, 2016. doi: 10.1103/PhysRevD.93.112009. https://arxiv.org/abs/1507.07129.

[232] CMS Collaboration. Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state. *Phys. Lett.*, B771:80–105, 2017. doi: 10.1016/j.physletb.2017.05.019. https://arxiv.org/abs/1612.00999.

[233] CMS Collaboration. Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton-proton collisions at sqrt(s) = 13 TeV. 2017. https://arxiv.org/abs/1708.01062.

[234] CMS Collaboration. Search for electroweak production of a vector-like quark decaying to a top quark and a Higgs boson using boosted topologies in fully hadronic final states. *JHEP*, 04:136, 2017. doi: 10.1007/JHEP04(2017)136. https://arxiv.org/abs/1612.05336.

[235] CMS Collaboration. Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 05:029, 2017. doi: 10.1007/JHEP05(2017)029. https://arxiv.org/abs/1701.07409.

[236] CMS Collaboration. Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:085, 2017. doi: 10.1007/JHEP11(2017)085. https://arxiv.org/abs/1706.03408.

[237] CMS Collaboration. Search for a heavy resonance decaying to a top quark and a vector-like top quark at $\sqrt{s} = 13$ TeV. *JHEP*, 09:053, 2017. doi: 10.1007/JHEP09(2017)053. https://arxiv.org/abs/1703.06352.

[238] CMS Collaboration. Search for pair production of vector-like quarks in the bW$\bar{b}$W channel from proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 779:82, 2018. doi: 10.1016/j.physletb.2018.01.077. https://arxiv.org/abs/1710.01539.

[239] CMS Collaboration. Search for vector-like light-flavor quark partners in proton-proton collisions at sqrt(s)=8 TeV. 2017. https://arxiv.org/abs/1708.02510.

[240] ATLAS Collaboration. Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 02:110, 2016. doi: 10.1007/JHEP02(2016)110. https://arxiv.org/abs/1510.02664.

[241] ATLAS Collaboration. Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J.*, C76(8):442, 2016. doi: 10.1140/epjc/s10052-016-4281-8. https://arxiv.org/abs/1602.05606.

[242] ATLAS Collaboration. Search for pair production of heavy vector-like quarks decaying to high-p$_T$ W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:141, 2017. doi: 10.1007/JHEP10(2017)141. https://arxiv.org/abs/1707.03347.

[243] ATLAS Collaboration. Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *JHEP*, 08:052, 2017. doi: 10.1007/JHEP08(2017)052. https://arxiv.org/abs/1705.10751.

[244] ATLAS Collaboration. Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018. https://arxiv.org/abs/1801.08769.

[245] CMS Collaboration. Searches for third-generation squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 06:116, 2015. doi: 10.1007/JHEP06(2015)116. https://arxiv.org/abs/1503.08037.

[246] CMS Collaboration. Search for supersymmetry in pp collisions at $\sqrt{s} = 8$ TeV in final states with boosted W bosons and b jets using razor variables. *Phys. Rev.*, D93(9):092009, 2016. doi: 10.1103/PhysRevD.93.092009. https://arxiv.org/abs/1602.02917.

[247] CMS Collaboration. Search for direct pair production of supersymmetric top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76(8):460, 2016. doi: 10.1140/epjc/s10052-016-4292-5. https://arxiv.org/abs/1603.00765.

[248] CMS Collaboration. Search for supersymmetry in the all-hadronic final state using top quark tagging in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev.*, D96(1):012004, 2017. doi: 10.1103/PhysRevD.96.012004. https://arxiv.org/abs/1701.01954.

[249] CMS Collaboration. Search for supersymmetry in proton-proton collisions at 13 TeV using identified

top quarks. *Phys. Rev.*, D97(1):012007, 2018. doi: 10.1103/PhysRevD.97.012007. https://arxiv.org/abs/1710.11188.

[250] CMS Collaboration. Search for physics beyond the standard model in events with high-momentum Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV. *Submitted to: Phys. Rev. Lett.*, 2017. https://arxiv.org/abs/1712.08501.

[251] CMS Collaboration. Search for dark matter in proton-proton collisions at 8 TeV with missing transverse momentum and vector boson tagged jets. *JHEP*, 12:083, 2016. doi: 10.1007/JHEP12(2016)083, 10.1007/JHEP08(2017)035. https://arxiv.org/abs/1607.05764. [Erratum: JHEP08,035(2017)].

[252] CMS Collaboration. Search for high-mass $Z\gamma$ resonances in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV using jet substructure techniques. *Phys. Lett.*, B772:363–387, 2017. doi: 10.1016/j.physletb.2017.06.062. https://arxiv.org/abs/1612.09516.

[253] CMS Collaboration. Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV. *JHEP*, 07:014, 2017. doi: 10.1007/JHEP07(2017)014. https://arxiv.org/abs/1703.05236.

[254] CMS Collaboration. Search for associated production of dark matter with a Higgs boson decaying to b$\bar{\text{b}}$ or $\gamma\gamma$ at $\sqrt{s} = 13$ TeV. *JHEP*, 10:180, 2017. doi: 10.1007/JHEP10(2017)180. https://arxiv.org/abs/1703.05236.

[255] CMS Collaboration. Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV. 2017. https://arxiv.org/abs/1712.02345.

[256] CMS Collaboration. Search for $Z\gamma$ resonances using leptonic and hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2017. https://arxiv.org/abs/1712.03143.

[257] CMS Collaboration. Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at $\sqrt{s} = 13$ TeV. 2018. https://arxiv.org/abs/1801.08427.

[258] CMS Collaboration. Search for $W' \to tb$ in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 02:122, 2016. doi: 10.1007/JHEP02(2016)122. https://arxiv.org/abs/1509.06051.

[259] CMS Collaboration. Search for the production of an excited bottom quark decaying to tW in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 01:166, 2016. doi: 10.1007/JHEP01(2016)166. https://arxiv.org/abs/1509.08141.

[260] CMS Collaboration. Searches for W' bosons decaying to a top quark and a bottom quark in proton-proton collisions at 13 TeV. *JHEP*, 08:029, 2017. doi: 10.1007/JHEP08(2017)029. https://arxiv.org/abs/1706.04260.

[261] CMS Collaboration. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}$ =13TeV. *Phys. Lett.*, B778:101–127, 2018. doi: 10.1016/j.physletb.2018.01.001. https://arxiv.org/abs/1707.02909.

[262] ATLAS Collaboration. Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev.*, D93(7):072007, 2016. doi: 10.1103/PhysRevD.93.072007. https://arxiv.org/abs/1510.06218.

[263] ATLAS Collaboration. Search for dark matter in events with a hadronically decaying W or Z boson and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 112(4):041802, 2014. doi: 10.1103/PhysRevLett.112.041802. https://arxiv.org/abs/1309.4017.

[264] ATLAS Collaboration. Search for dark matter produced in association with a hadronically decaying vector boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B763:251–268, 2016. doi: 10.1016/j.physletb.2016.10.042. https://arxiv.org/abs/1608.02372.

[265] ATLAS Collaboration. Search for heavy resonances decaying to a Z boson and a photon in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B764:11–30, 2017. doi: 10.1016/j.physletb.2016.11.005. https://arxiv.org/abs/1607.06363.

[266] ATLAS Collaboration. A search for resonances decaying into a Higgs boson and a new particle X in the $XH \rightarrow qqbb$ final state with the ATLAS detector. *Phys. Lett.*, B779:24–45, 2018. doi: 10.1016/j.physletb.2018.01.042. https://arxiv.org/abs/1709.06783.

[267] ATLAS Collaboration. Search for $W' \rightarrow tb \rightarrow qqbb$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J.*, C75(4):165, 2015. doi: 10.1140/epjc/s10052-015-3372-2. https://arxiv.org/abs/1408.0886.

[268] ATLAS Collaboration. Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s}$=13 TeV pp collision data with the ATLAS detector. 2017. https://arxiv.org/abs/1712.02332.

[269] ATLAS Collaboration. Search for Supersymmetry in final states with missing transverse momentum and multiple b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2017. https://arxiv.org/abs/1711.01901.

[270] ATLAS Collaboration. Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s}$=13 TeV with the ATLAS detector. *JHEP*, 12:085, 2017. doi: 10.1007/JHEP12(2017)085. https://arxiv.org/abs/1709.04183.

[271] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012. doi: 10.1016/j.physrep.2012.02.002. https://arxiv.org/abs/1106.0034.

[272] Christopher T. Hill. Topcolor assisted technicolor. *Phys. Lett.*, B345:483–489, 1995. doi: 10.1016/0370-2693(94)01660-5. https://arxiv.org/abs/hep-ph/9411426.

[273] David B. Kaplan and Howard Georgi. SU(2) x U(1) Breaking by Vacuum Misalignment. *Phys. Lett.*, B136:183–186, 1984. doi: 10.1016/0370-2693(84)91177-8. https://inspirehep.net/record/192986.

[274] David B. Kaplan, Howard Georgi, and Savas Dimopoulos. Composite Higgs Scalars. *Phys. Lett.*, B136:187–190, 1984. doi: 10.1016/0370-2693(84)91178-X. https://inspirehep.net/record/193935.

[275] Howard Georgi, David B. Kaplan, and Peter Galison. Calculation of the Composite Higgs Mass. *Phys. Lett.*, B143:152–154, 1984. doi: 10.1016/0370-2693(84)90823-2. https://inspirehep.net/record/14863.

[276] Tom Banks. Constraints on SU(2) x U(1) breaking by vacuum misalignment. *Nucl. Phys.*, B243:125–130, 1984. doi: 10.1016/0550-3213(84)90389-4. https://inspirehep.net/record/199825.

[277] Howard Georgi and David B. Kaplan. Composite Higgs and Custodial SU(2). *Phys. Lett.*, B145:216–220, 1984. doi: 10.1016/0370-2693(84)90341-1. https://inspirehep.net/record/203277.

[278] Michael J. Dugan, Howard Georgi, and David B. Kaplan. Anatomy of a Composite Higgs Model. *Nucl. Phys.*, B254:299–326, 1985. doi: 10.1016/0550-3213(85)90221-4. https://inspirehep.net/record/205792.

[279] Howard Georgi. A Tool Kit for Builders of Composite Models. *Nucl. Phys.*, B266:274–284, 1986. doi: 10.1016/0550-3213(86)90092-1. https://inspirehep.net/record/216843.

[280] Brando Bellazzini, Csaba Csáki, and Javi Serra. Composite Higgses. *Eur. Phys. J.*, C74(5):2766, 2014. doi: 10.1140/epjc/s10052-014-2766-x. https://arxiv.org/abs/1401.2457.

[281] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999. doi: 10.1103/PhysRevLett.83.3370. http://arxiv.org/abs/hep-ph/9905221.

[282] Kaustubh Agashe, Antonio Delgado, Michael J. May, and Raman Sundrum. RS1, custodial isospin and precision tests. *JHEP*, 08:050, 2003. doi: 10.1088/1126-6708/2003/08/050. http://arxiv.org/abs/hep-ph/0308036.

[283] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. Bulk gauge fields in the Randall-Sundrum model. *Phys. Lett.*, B473:43–49, 2000. doi: 10.1016/S0370-2693(99)01430-6. http://arxiv.org/abs/hep-ph/9911262.

[284] Alex Pomarol. Gauge bosons in a five-dimensional theory with localized gravity. *Phys. Lett.*, B486:153–157, 2000. doi: 10.1016/S0370-2693(00)00737-1. http://arxiv.org/abs/hep-ph/9911294.

[285] Keith Rehermann and Brock Tweedie. Efficient Identification of Boosted Semileptonic Top Quarks at the LHC. *JHEP*, 03:059, 2011. doi: 10.1007/JHEP03(2011)059. http://arxiv.org/abs/arXiv:1007.2221.

[286] Duccio Pappadopulo, Andrea Thamm, Riccardo Torre, and Andrea Wulzer. Heavy Vector Triplets: Bridging Theory and Data. *JHEP*, 09:060, 2014. doi: 10.1007/JHEP09(2014)060. http://arxiv.org/abs/arXiv:1402.4431.

[287] Nima Arkani-Hamed, Andrew G. Cohen, and Howard Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Phys. Lett. B*, 513:232, 2001. doi: 10.1016/S0370-2693(01)00741-9. http://arxiv.org/abs/hep-ph/0105239.

[288] Martin Schmaltz and David Tucker-Smith. Little Higgs review. *Ann. Rev. Nucl. Part. Sci.*, 55:229, 2005. doi: 10.1146/annurev.nucl.55.090704.151502. http://arxiv.org/abs/hep-ph/0502182.

[289] Ignatios Antoniadis, K. Benakli, and M. Quiros. Finite Higgs mass without supersymmetry. *New J. Phys.*, 3:20, 2001. doi: 10.1088/1367-2630/3/1/320. http://arxiv.org/abs/hep-th/0108005.

[290] Yutaka Hosotani, Shusaku Noda, and Kazunori Takenaga. Dynamical gauge-Higgs unification in the electroweak theory. *Phys. Lett. B*, 607:276, 2005. doi: 10.1016/j.physletb.2004.12.029. http://arxiv.org/abs/hep-ph/0410193.

[291] Kaustubh Agashe, Roberto Contino, and Alex Pomarol. The minimal composite Higgs model. *Nucl. Phys. B*, 719:165, 2005. doi: 10.1016/j.nuclphysb.2005.04.035. http://arxiv.org/abs/hep-ph/0412089.

[292] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev. D*, 88 (9):094010, 2013. doi: 10.1103/PhysRevD.88.094010. http://arxiv.org/abs/arXiv:1306.0572.

[293] U. Baur, I. Hinchliffe, and D. Zeppenfeld. Excited Quark Production at Hadron Colliders. *Int. J. Mod. Phys.*, A2:1285, 1987. doi: 10.1142/S0217751X87000661. https://inspirehep.net/record/247106.

[294] JoAnne L. Hewett and Thomas G. Rizzo. Low-Energy Phenomenology of Superstring Inspired E(6) Models. *Phys. Rept.*, 183:193, 1989. doi: 10.1016/0370-1573(89)90071-9. https://inspirehep.net/record/268529.

[295] U. Baur, M. Spira, and P. M. Zerwas. Excited Quark and Lepton Production at Hadron Colliders. *Phys. Rev.*, D42:815–824, 1990. doi: 10.1103/PhysRevD.42.815. https://inspirehep.net/record/283232.

[296] Paul Langacker. The Physics of Heavy Z' Gauge Bosons. *Rev. Mod. Phys.*, 81:1199–1228, 2009. doi: 10.1103/RevModPhys.81.1199. https://arxiv.org/abs/0801.1345.

[297] Haipeng An, Ran Huo, and Lian-Tao Wang. Searching for Low Mass Dark Portal at the LHC. *Phys. Dark Univ.*, 2:50–57, 2013. doi: 10.1016/j.dark.2013.03.002. https://arxiv.org/abs/1212.2221.

[298] Arvind Rajaraman, William Shepherd, Tim M. P. Tait, and Alexander M. Wijangco. LHC Bounds on Interactions of Dark Matter. *Phys. Rev.*, D84:095013, 2011. doi: 10.1103/PhysRevD.84.095013. https://arxiv.org/abs/1108.1196.

[299] Jessica Goodman, Masahiro Ibe, Arvind Rajaraman, William Shepherd, Tim M. P. Tait, and Hai-Bo Yu. Constraints on Dark Matter from Colliders. *Phys. Rev.*, D82:116010, 2010. doi: 10.1103/PhysRevD.82.116010. https://arxiv.org/abs/1008.1783.

[300] Keisuke Fujii et al. Physics Case for the International Linear Collider. 2015. http://arxiv.org/abs/arXiv:1506.05992.

[301] Howard Baer, Tim Barklow, Keisuke Fujii, Yuanning Gao, Andre Hoang, et al. The International Linear Collider Technical Design Report - Volume 2: Physics. 2013.

[302] M J Boland et al. Updated baseline for a staged Compact Linear Collider. 2016. doi: 10.5170/CERN-2016-004. http://arxiv.org/abs/arXiv:1608.07537.

[303] M Aicheler, P Burrows, M Draper, T Garvey, P Lebrun, K Peach, N Phinney, H Schmickler, D Schulte, and N Toge. A Multi-TeV Linear Collider Based on CLIC Technology. 2012. doi: 10.5170/CERN-2012-007. https://inspirehep.net/record/1228241.

[304] M. Bicer, H. Duran Yildiz, I. Yildiz, G. Coignet, M. Delmastro, et al. First Look at the Physics Case of TLEP. *JHEP*, 01:164, 2014. doi: 10.1007/JHEP01(2014)164. http://arxiv.org/abs/arXiv:1308.6176.

[305] CEPC-SPPC Study Group. CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector, IHEP-CEPC-DR-2015-01 (2015). 2015.

[306] Patrick Janot. Top-quark electroweak couplings at the FCC-ee. *JHEP*, 04:182, 2015. doi: 10.1007/JHEP04(2015)182. http://arxiv.org/abs/arXiv:1503.01325.

[307] T. Barklow, J. Brau, K. Fujii, J. Gao, J. List, N. Walker, and K. Yokoya. ILC Operating Scenarios. 2015. https://arxiv.org/abs/1506.07830.

[308] Nima Arkani-Hamed, Tao Han, Michelangelo Mangano, and Lian-Tao Wang. Physics Opportunities of a 100 TeV Proton-Proton Collider. *Phys. Rept.*, 652:1–49, 2016. doi: 10.1016/j.physrep.2016.07.004. http://arxiv.org/abs/arXiv:1511.06495.

[309] T. Golling et al. Physics at a 100 TeV pp collider: beyond the Standard Model phenomena. *CERN Yellow Report*, (3):441–634, 2017. doi: 10.23731/CYRM-2017-003.441. http://arxiv.org/abs/arXiv:1606.00947.

[310] R. Contino et al. Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies. *CERN Yellow Report*, (3):255–440, 2017. doi: 10.23731/CYRM-2017-003.255. http://arxiv.org/abs/arXiv:1606.09408.

[311] M. L. Mangano et al. Physics at a 100 TeV pp collider: Standard Model processes. *CERN Yellow Report*, (3):1–254, 2017. doi: 10.23731/CYRM-2017-003.1. http://arxiv.org/abs/arXiv:1607.01831.

[312] Ian Hinchliffe, Ashutosh Kotwal, Michelangelo L. Mangano, Chris Quigg, and Lian-Tao Wang. Luminosity goals for a 100-TeV pp collider. *Int. J. Mod. Phys.*, A30(23):1544002, 2015. doi: 10.1142/S0217751X15440029. http://arxiv.org/abs/arXiv:1504.06108.

[313] Michelangelo L. Mangano, Tilman Plehn, Peter Reimitz, Torben Schell, and Hua-Sheng Shao. Measuring the Top Yukawa Coupling at 100 TeV. *J. Phys.*,

[314] Alan J. Barr, Matthew J. Dolan, Christoph Englert, Danilo Enoque Ferreira de Lima, and Michael Spannowsky. Higgs Self-Coupling Measurements at a 100 TeV Hadron Collider. *JHEP*, 02:016, 2015. doi: 10.1007/JHEP02(2015)016. http://arxiv.org/abs/arXiv:1412.7154.

[315] Marco Farina, Giuliano Panico, Duccio Pappadopulo, Joshua T. Ruderman, Riccardo Torre, and Andrea Wulzer. Energy helps accuracy: electroweak precision tests at hadron colliders. *Phyis. Lett.*, B772:210–215, 2017. doi: 10.1016/j.physletb.2017.06.043. http://arxiv.org/abs/arXiv:1609.08157.

[316] Juan A. Aguilar-Saavedra, Benjamin Fuks, and Michelangelo L. Mangano. Pinning down top dipole moments with ultra-boosted tops. *Phys. Rev.*, D91:094021, 2015. doi: 10.1103/PhysRevD.91.094021. http://arxiv.org/abs/arXiv:1412.6654.

[317] Christoph Englert, Karl Nordstrom, Liam Moore, and Michael Russell. Giving top quark effective operators a boost. *Phys. Lett.*, B763:9–15, 2016. doi: 10.1016/j.physletb.2016.10.021. http://arxiv.org/abs/arXiv:1607.04304.

[318] Martin Perello Rosello and Marcel Vos. Constraints on four-fermion interactions from the $t\bar{t}$ charge asymmetry at hadron colliders. *Eur. Phys. J.*, C76(4):200, 2015. doi: 10.1140/epjc/s10052-016-4040-x. http://arxiv.org/abs/arXiv:1512.07542.

[319] S. V. Chekanov, M. Beydler, A. V. Kotwal, L. Gray, S. Sen, N. V. Tran, S. S. Yu, and J. Zuzelski. Initial performance studies of a general-purpose detector for multi-TeV physics at a 100 TeV pp collider. *JINST*, 12(06):P06009, 2017. doi: 10.1088/1748-0221/12/06/P06009. https://arxiv.org/abs/1612.07291.

[320] Shin-Shan Yu, Sergei Chekanov, Lindsey Gray, Ashutosh Kotwal, Sourav Sen, and Nhan Viet Tran. Study Of Boosted W-Jets And Higgs-Jets With the SiFCC Detector. In *38th International Conference on High Energy Physics (ICHEP 2016) Chicago, IL, USA, August 03-10, 2016*, 2016. http://inspirehep.net/record/1495717/files/ arXiv:1611.01136.pdf.

[321] T. Carli, C. Helsens, A. Henriques Correia, and C. Solans Sanchez. Containment and resolution of

hadronic showers at the FCC. *JINST*, 11(09):P09012, 2016. doi: 10.1088/1748-0221/11/09/P09012. http://arxiv.org/abs/arXiv:1604.01415.

[322] Andrew J. Larkoski, Fabio Maltoni, and Michele Selvaggi. Tracking down hyper-boosted top quarks. *JHEP*, 06:032, 2015. doi: 10.1007/JHEP06(2015)032. http://arxiv.org/abs/arXiv:1503.03347.

[323] Michelangelo Mangano. *Physics at the FCC-hh, a 100 TeV pp collider*. CERN Yellow Reports: Monographs. CERN, Geneva, 2017. http://cds.cern.ch/record/2270978.

[324] H. Abramowicz et al. Higgs physics at the CLIC electron–positron linear collider. *Eur. Phys. J.*, C77 (7):475, 2017. doi: 10.1140/epjc/s10052-017-4968-5. https://arxiv.org/abs/1608.07538.

[325] M. S. Amjad et al. A precise characterisation of the top quark electro-weak vertices at the ILC. *Eur. Phys. J.*, C75(10):512, 2015. doi: 10.1140/epjc/s10052-015-3746-5. http://arxiv.org/abs/arXiv:1505.06020.

[326] Mark Thomson. Model-independent measurement of the e$^+$ e$^-$ → HZ cross section at a future e$^+$ e$^-$ linear collider using hadronic Z decays. *Eur. Phys. J.*, C76(2):72, 2016. doi: 10.1140/epjc/s10052-016-3911-5. https://arxiv.org/abs/1509.02853.

[327] M. Boronat, J. Fuster, Ignacio Garcia, E. Ros, and Marcel Vos. A robust jet reconstruction algorithm for high-energy lepton colliders. *Phys. Lett.*, B750:95–99, 2015. doi: 10.1016/j.physletb.2015.08.055. https://arxiv.org/abs/1404.4294.

[328] M. Boronat, J. Fuster, I. Garcia, Ph. Roloff, R. Simoniello, and M. Vos. Jet reconstruction at high-energy lepton colliders. 2016. https://arxiv.org/abs/1607.05039.

[329] Javier Aparisi, Ignacio García, Martin Perelló, Ph Roloff, Rosa Simoniello, and Marcel Vos. Jet reconstruction algorithms in e^+e collisions. In *Parton Radiation and Fragmentation from LHC to FCC-ee*, pages 128–133, 2017. https://inspirehep.net/record/1513006/files/ 1512294_128-133.pdf.

[330] Halina Abramowicz et al. The International Linear Collider Technical Design Report - Volume 4: Detectors. 2013. http://arxiv.org/abs/arXiv:1306.6329.

[331] C. Adloff et al. Construction and Commissioning of the CALICE Analog Hadron Calorimeter Prototype. *JINST*, 5:P05004, 2010. doi: 10.1088/1748-0221/5/05/P05004. https://arxiv.org/abs/1003.2662.

[332] C. Adloff et al. Response of the CALICE Si-W electromagnetic calorimeter physics prototype to electrons. *Nucl. Instrum. Meth.*, A608:372–383, 2009. doi: 10.1016/j.nima.2009.07.026. https://arxiv.org/abs/0811.2354.

[333] J. S. Marshall and M. A. Thomson. The Pandora Software Development Kit for Pattern Recognition. *Eur. Phys. J.*, C75(9):439, 2015. doi: 10.1140/epjc/s10052-015-3659-3. https://arxiv.org/abs/1506.05348.

[334] David d'Enterria, editor. *Parton Radiation and Fragmentation from LHC to FCC-ee*, 2017. http://inspirehep.net/record/1512294/files/arXiv:1702.01329.pdf.

[335] J. S. Marshall, A. Münnich, and M. A. Thomson. Performance of Particle Flow Calorimetry at CLIC. *Nucl. Instrum. Meth.*, A700:153–162, 2013. doi: 10.1016/j.nima.2012.10.038. https://arxiv.org/abs/1209.4039.

[336] Boost 2016 Participants. Boost Zurich 2016. https://indico.cern.ch/event/439039/.

[337] Boost 2017 Participants. Boost Buffalo 2017. https://indico.cern.ch/event/579660/.