

# Particle-level pileup subtraction for jets and jet shapes

Peter Berta,<sup>a</sup> Martin Spousta,<sup>a</sup> David W. Miller,<sup>b</sup> and Rupert Leitner<sup>a</sup>

<sup>a</sup>*Faculty of Mathematics and Physics, Charles University in Prague,  
V Holešovičkách 2, Prague, Czech Republic*

<sup>b</sup>*Enrico Fermi Institute, University of Chicago,  
Chicago, IL, U.S.A.*

*E-mail:* [berta@ipnp.troja.mff.cuni.cz](mailto:berta@ipnp.troja.mff.cuni.cz), [Martin.Spousta@cern.ch](mailto:Martin.Spousta@cern.ch),

[David.W.Miller@uchicago.edu](mailto:David.W.Miller@uchicago.edu), [Rupert.Leitner@mff.cuni.cz](mailto:Rupert.Leitner@mff.cuni.cz)

**ABSTRACT:** We present an extension to the jet area-based pileup subtraction for both jet kinematics and jet shapes. A particle-level approach is explored whereby the jet constituents are corrected or removed using an extension of the methods currently being employed by the LHC experiments. Several jet shapes and nominal jet radii are used to assess the performance in simulated events with pileup levels equivalent to approximately 30 and 100 interactions per bunch crossing, which are characteristic of both the LHC Run I and Run II conditions. An improved performance in removing the pileup contributions is found when using the new subtraction method. The performance of the new procedure is also compared to other existing methods.

**KEYWORDS:** Jet substructure, Jets, Hadron-Hadron Scattering, Event-by-event fluctuation

ARXIV EPRINT: [1403.3108](https://arxiv.org/abs/1403.3108)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Subtraction algorithm</b>	<b>2</b>
<b>3</b>	<b>Performance of the subtraction</b>	<b>5</b>
3.1	Jet kinematics	6
3.2	Jet shape definitions	8
3.3	Jet shape subtraction	9
<b>4</b>	<b>Conclusions</b>	<b>12</b>

---

## 1 Introduction

The intense environment created by the high luminosity of the Large Hadron Collider (LHC) necessitates novel methods for isolating, mitigating, and, where possible, correcting for the contributions of multiple uncorrelated proton-proton interactions (pileup) to the measured hadronic final state. Pileup has a substantial impact on both jet kinematics and substructure, thereby degrading critical tools for identifying new physics via highly boosted hadronic decays of  $W$ ,  $Z$ , and Higgs bosons, or top quarks.

The last few years has seen the development of several effective tools for pileup mitigation and removal. Simple subtraction techniques remove a constant offset from the measured transverse momentum that is proportional to the number of observed pileup events [1]. So-called *grooming techniques* such as filtering [2], pruning [3], and trimming [4], actively remove potential pileup constituents from jets. Other approaches, such as the *jet cleansing method* [5] or *charged hadron subtraction* [6], use tracking information to identify a given hadronic energy deposition with charged particles originating in pileup interactions.

Techniques that utilize event-by-event and jet-by-jet information to determine the extent of contamination from pileup provide a new approach to perform jet physics at very high luminosities. The *area-based subtraction* procedure [7] corrects the jet 4-momentum and it is extended to account for hadron masses in ref. [8]. The *shape-expansion method* [8] provides general approach to correct jet shapes.

The extension of the area-based subtraction procedure that we propose here allows for a particle-by-particle approach to this concept. We find improved performance in removing the contributions due to pileup using this new procedure even for previously intractable jet shape observables (such as planar flow). Furthermore, this approach provides the possibility to perform pileup subtraction without explicit consideration of a specific jet algorithm, reducing the constraints and increasing the flexibility of the jet area-based subtraction procedure overall. Therefore, this method may be used also in heavy ion physics where the jet reconstruction is challenging due to sizable underlying event [9–11].

## 2 Subtraction algorithm

The novel feature of the approach described here is the *local* subtraction of pileup at the level of individual jet constituents. In contrast to the area-based subtraction and the shape-expansion method, the constituent-level subtraction is performed particle-by-particle, thereby correcting both the 4-momentum of the jet and its substructure, simultaneously. This is achieved by combining the kinematics of particles within a specific jet with the kinematics of soft “negative” particles that are added to balance the pileup contribution.

The basic ingredient of the particle-level subtraction is the pileup energy density estimation which is identical to that used in the shape-expansion method proposed in ref. [8]. The contamination due to pileup is described in terms of the transverse momentum density  $\rho$  and mass density  $\rho_m$ . The expected pileup deposition in a small region of  $\Delta y \Delta \phi$  is expressed by the 4-momentum

$$p_{\text{pileup}}^\mu = [\rho \cos \phi, \rho \sin \phi, (\rho + \rho_m) \sinh y, (\rho + \rho_m) \cosh y] \cdot \Delta y \Delta \phi, \quad (2.1)$$

where the pileup  $p_T$  and mass densities,  $\rho$  and  $\rho_m$ , are assumed to be weakly dependent on rapidity  $y$  and azimuth  $\phi$ . In the shape-expansion method, all particles (or effective particles such as calorimeter towers [13] or clusters [14]) in the event are grouped into *patches* in order to estimate the densities used in eq. (2.1). The patches are defined by jets reconstructed using the  $k_t$  algorithm [15, 16]. The transverse momentum,  $p_{T\text{patch}}$ , and mass,  $m_{\delta\text{patch}}$ , of each patch is determined by summing over all particles within that patch:

$$p_{T\text{patch}} = \sum_{i \in \text{patch}} p_{Ti}, \quad m_{\delta\text{patch}} = \sum_{i \in \text{patch}} \left( \sqrt{m_i^2 + p_{Ti}^2} - p_{Ti} \right), \quad (2.2)$$

where  $p_{Ti}$  and  $m_i$  are the transverse momentum and mass of particle  $i$ , respectively. Each patch covers certain area  $A_{\text{patch}}$  in the  $(y - \phi)$  plane. The overall background  $p_T$  and mass densities are estimated as

$$\rho = \text{median}_{\text{patches}} \left\{ \frac{p_{T\text{patch}}}{A_{\text{patch}}} \right\}, \quad \rho_m = \text{median}_{\text{patches}} \left\{ \frac{m_{\delta\text{patch}}}{A_{\text{patch}}} \right\}, \quad (2.3)$$

although several modifications exist, including  $y$ -dependent  $\rho$  and  $\rho_m$  [17]. The estimation of the background densities is the first step which needs to be followed by a scheme by which to subtract a specified amount of those densities.

In our approach, massless particles with very low momentum are incorporated into the event such that they uniformly cover the  $y - \phi$  plane with high density. These soft particles are referred to as *ghosts* and they are most commonly used to define the area of a jet [12] or to perform the shape-expansion correction. Each ghost covers a certain fixed area,  $A_g$ , in the  $y - \phi$  plane which is defined by the ghost number density ( $A_g$  is its inverse). The 4-momentum of each particle or ghost is expressed as

$$p^\mu = [p_T \cos \phi, p_T \sin \phi, (p_T + m_\delta) \sinh y, (p_T + m_\delta) \cosh y], \quad (2.4)$$

where  $m_\delta = \sqrt{m^2 + p_T^2} - p_T$  (in what follows, we will use superscript  $g$  to denote the kinematic variables of ghosts). After adding ghosts into the event, the jet clustering algorithm runs over all particles and ghosts delivering the same jets as in the case without the ghosts. Now, the jets contain except the real particles also ghosts which can be used to correct for the pileup contribution within each jet. Eq. (2.1) is translated into the 4-momentum of each ghost by identifying the transverse momentum  $p_T^g$  and mass  $m_\delta^g$  with the amount of pileup within area  $A_g$ :

$$\begin{aligned} p_T^g &= A_g \cdot \rho, \\ m_\delta^g &= A_g \cdot \rho_m. \end{aligned} \tag{2.5}$$

An iterative procedure is used to define the scheme for calculating the specified amount of transverse momentum and mass  $m_\delta$  to subtract from each jet constituent. For each pair of particle  $i$  and ghost  $k$ , a matching scheme is implemented using the distance measure,  $\Delta R_{i,k}$ , defined as

$$\Delta R_{i,k} = p_{Ti}^\alpha \cdot \sqrt{(y_i - y_k^g)^2 + (\phi_i - \phi_k^g)^2}. \tag{2.6}$$

For complete generality,  $\alpha$  is allowed to be any real number, but is taken to be zero in the studies performed here. The list of all distance measures,  $\{\Delta R_{i,k}\}$ , is sorted from the lowest to the highest values. The pileup removal proceeds iteratively, starting from the particle-ghost pair with the lowest  $\Delta R_{i,k}$ . At each step, the momentum  $p_T$  and mass  $m_\delta$  of each particle  $i$  and ghost  $k$  are modified as follows.

$$\left. \begin{array}{l} \text{If } p_{Ti} \geq p_{Tk}^g : \\ \quad p_{Ti} \longrightarrow p_{Ti} - p_{Tk}^g, \\ \quad p_{Tk}^g \longrightarrow 0; \\ \text{otherwise:} \\ \quad p_{Ti} \longrightarrow 0, \\ \quad p_{Tk}^g \longrightarrow p_{Tk}^g - p_{Ti}. \end{array} \right| \begin{array}{l} \text{If } m_{\delta i} \geq m_{\delta k}^g : \\ \quad m_{\delta i} \longrightarrow m_{\delta i} - m_{\delta k}^g, \\ \quad m_{\delta k}^g \longrightarrow 0; \\ \text{otherwise:} \\ \quad m_{\delta i} \longrightarrow 0, \\ \quad m_{\delta k}^g \longrightarrow m_{\delta k}^g - m_{\delta i}. \end{array} \tag{2.7}$$

The azimuth and rapidity of the particles and ghosts remain unchanged. The iterative process is terminated when the end of the sorted list is reached. Alternatively, a threshold  $\Delta R^{\max}$  can be introduced to stop the iterations when  $\Delta R_{i,k} > \Delta R^{\max}$ . In principle, introducing the  $\Delta R^{\max}$  threshold also guarantees that only ghosts neighbouring a given particle are used to correct the kinematics of that particle. Particles with zero transverse momentum after the iterative process are discarded and the 4-momentum of a given jet is recalculated following a desired recombination scheme (commonly, the 4-momentum recombination scheme is used [18]). It can happen that after the subtraction no real particle remains. This may be a signal that such a jet originates from pileup.

The scalar subtraction in eq. (2.7) is chosen instead of 4-momentum subtraction since it allows to reduce the local differences between the actual background deposit and its estimate from eq. (2.1).<sup>1</sup> The scalar subtraction also eliminates the occurrence of unphysical

---

<sup>1</sup>An alternative form of eq. (2.7) with subtracting the 4-momenta would lead to a corrected jet with the same 4-momentum as if the area-based subtraction from ref. [8] was applied which follows from the additivity of 4-momenta. The performance of such particle-level correction will be explored in an upcoming study.

negative squared masses which may be present when subtracting 4-momenta. Furthermore, the scalar summation is also used in the calculation of the transverse momentum of patches in eq. (2.2) and thereby it avoids the dependence of background densities in eq. (2.3) on the shape and size of the patches.

The above described subtraction procedure — referred to as *constituent subtraction* — corrects both the 4-momentum of a jet as well as its substructure. The constituent subtraction works equally well when applied directly to Monte Carlo truth particles from simulation as when applied to a coarse pseudo-detector grid over which the energy from the truth particles is distributed (see section 3.3). An important feature of the algorithm is that it preserves longitudinal invariance — an arbitrary jet after the correction and a subsequent boost in the direction of colliding beams has the same constituents as the same jet which is first boosted and then corrected. The algorithm can be easily extended to account for rapidity dependence of background densities  $\rho$  and  $\rho_m$  in eq. (2.5).

It is also straightforward to extend the method further to the whole event instead of correcting just particles within a jet. Jet finding can then be performed using the subtracted event. Global event shapes can also benefit from the correction in addition to individual jet observables. The performance and resolution of missing transverse energy calculated from calorimeter energy deposits may also improve, thereby enhancing several searches for physics beyond the Standard Model. Studies of whole-event constituent subtraction are left to future work, as well as testing directly within the experimental communities.

An important advantage of the constituent subtraction is the speed — it can be as much as twenty times faster compared to the shape-expansion method, depending on the type of the jet shape and the nominal jet radius. Furthermore, the shape-expansion correction must be determined for each jet shape in consideration, whereas the constituent subtraction approach provides a corrected set of constituents, from which any shape may be determined. Corrected constituents may also then be used as inputs to jet grooming and tagging algorithms, e.g. the top-quark tagging using the shower deconstruction method [19]. In comparison to the jet cleansing method or charged hadron subtraction used by the CMS experiment, the constituent subtraction does not require any knowledge about the connection of each charged particle with the signal vertex or pileup vertices, though such a knowledge might in principle be used to further enhance capabilities of the algorithm.

The constituent subtraction procedure has following free parameters:  $A_g$ ,  $\Delta R^{\max}$ , and  $\alpha$ . The basic recommended settings are:  $A_g = 0.01$ ,  $\Delta R^{\max} \rightarrow \infty$ , and  $\alpha = 0$ . These settings were used in the performance studies presented in section 3. The subtraction is stable with respect to varying  $A_g$ . The variation of  $A_g$  by a factor of two does not lead to a change in any of the studied quantities that would be significant with respect to the statistical uncertainty shown on plots in section 3. Introducing a finite  $\Delta R^{\max}$  may improve the performance of the correction and the speed of the algorithm when running over the full event. The configuration with  $\alpha > 0$  prefers to subtract the lower  $p_T$  constituents which more often originate from background. This configuration will not be discussed in this paper but it appears to lead to an improvement in the correction of some of the jet shapes.

The software for the constituent subtraction is implemented as a part of the `FastJet Contrib` project [20].

### 3 Performance of the subtraction

The constituent subtraction algorithm corrects both the jet kinematics ( $p_T$  and mass) and the jet shapes. Studies of the algorithm performance for  $p_T$  are discussed in section 3.1 along with comparisons to the area-based subtraction which follows ref. [8] where the pileup 4-momentum (2.1) is subtracted using the jet area 4-vector  $A^\mu$  as

$$p_{\text{corr}}^\mu = [p^x - \rho A^x, p^y - \rho A^y, p^z - (\rho + \rho_m) A^y, E - (\rho + \rho_m) A^E]. \quad (3.1)$$

The performance of the subtraction applied to both the jet mass and several jet shapes is presented in section 3.3. Comparisons of the constituent subtraction approach with the shape-expansion method [8] are also presented.

The studies presented are performed using events generated with PYTHIA 8.180, tune 4C [21, 22] but without any detector simulation. The effect of additional proton-proton collisions is emulated by using inclusive events (often referred to as “minimum bias” events) overlaid with the hard scattering interaction, which are also generated with PYTHIA 8.180. The CTEQ 5L, LO parton density functions [23], configured to simulate the LHC conditions at  $\sqrt{s} = 8$  TeV, are used for all event generation. Two processes are simulated without underlying event: di-jet events covering the  $p_T$  range of 10-800 GeV and events with boosted top quarks from decay  $Z' \rightarrow t\bar{t}$  of hypothetical boson  $Z'$  with mass of 1.5 TeV. The performance of the subtraction is tested using jets clustered with the anti- $k_t$  algorithm [24] with the distance parameter  $R = 0.7$  or  $R = 1.0$  and jets clustered with Cambridge-Aachen (C/A) algorithm [25] with  $R = 1.2$ . These are representative jet definitions for both the ATLAS [26] and CMS [27] experimental collaborations. The number of pileup events,  $n_{PU}$ , has a Poisson distribution with a mean  $\langle n_{PU} \rangle$ . Two pileup conditions are simulated,  $\langle n_{PU} \rangle = 30$  and  $\langle n_{PU} \rangle = 100$ . These pileup configurations represent realistic conditions for LHC Run I and upcoming LHC Run II<sup>2</sup> as well as for the high luminosity LHC running [28]. On average, the pileup contribution to the hard-scatter event can be described through mean value of transverse momentum densities,  $\langle \rho \rangle$ , and pileup fluctuations characterized by standard deviation,  $\sigma[\rho]$ . For the used configuration  $\langle n_{PU} \rangle = 100$ , these quantities are  $\langle \rho \rangle \approx 75$  GeV and  $\sigma[\rho] \approx 13$  GeV.

All jet finding and background estimation is performed using `FastJet` 3.0.6 [17, 29]. The shape-expansion correction is performed using `FastJet Contrib` 1.003 [20]. The patches in eq. (2.3) are obtained by clustering particles with the  $k_t$  algorithm [15, 16] with distance parameter  $R = 0.4$ . The non-negligible dependence of the background densities  $\rho$  and  $\rho_m$  on rapidity impacts each of the corrections methods discussed below. Consequently,

---

<sup>2</sup>For LHC Run II, collisions at  $\sqrt{s} = 14$  TeV are planned. MC studies predict  $\langle n_{PU} \rangle$  of 120 and no significant increase in the size of pileup densities and inter-event fluctuations for this centre-of-mass energy.

in order to focus the comparisons and performance evaluations, only patches with rapidity  $|y| < 2.0$  are used in eq. (2.3) and jets are required to be fairly central, with  $|\eta| < 2.0$ .<sup>3</sup>

### 3.1 Jet kinematics

The ability of the subtraction to correctly recover the kinematics of the jet can be characterized in terms of following quantities: jet momentum response, jet momentum resolution, jet position resolution, and jet finding efficiency. These quantities are commonly used to evaluate the performance of the jet reconstruction, see e.g. refs. [30–33]. In this section we evaluate these quantities both for the constituent subtraction and area-based subtraction.

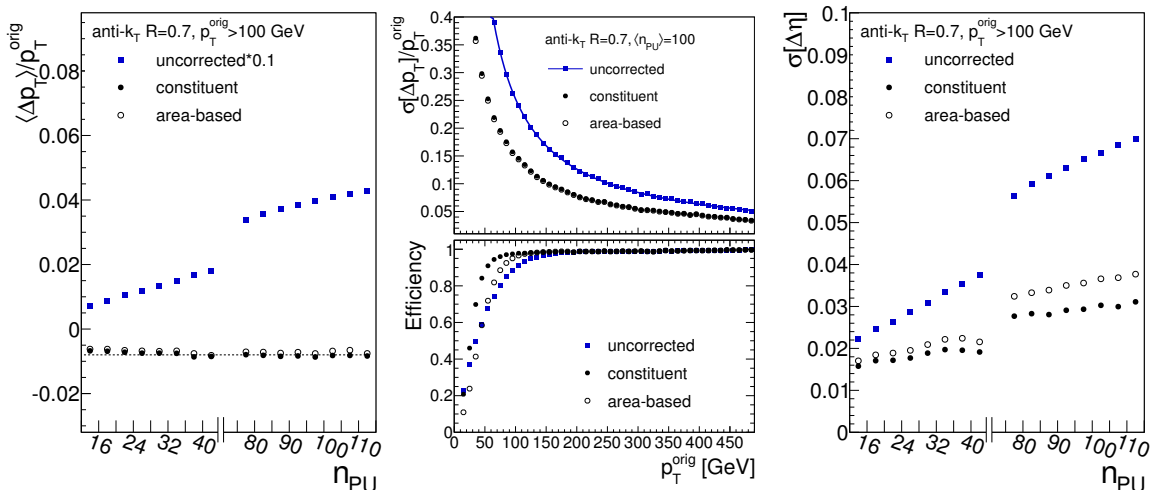
The first quantity characterizing the basic performance of the subtraction is the jet momentum response. It can be defined as  $\langle \Delta p_T \rangle / p_T^{\text{orig}} = \langle p_T - p_T^{\text{orig}} \rangle / p_T^{\text{orig}}$  where  $p_T^{\text{orig}}$  is the original jet momentum with no pileup and  $p_T$  is either the pileup corrected jet momentum or the momentum of uncorrected jet that is jet clustered in the presence of pileup with no subtraction. This quantity is also often referred to as the jet energy scale. In the optimal situation, the jet momentum response should be zero which means that, on average, the algorithm can reconstruct the same  $p_T$  as with no pileup. Left panel of figure 1 shows the jet momentum response as a function of the number of pileup collisions,  $n_{PU}$ , that is the size of the pileup. The jet momentum response of subtracted jets differs from zero by less than a 1%. The jet momentum response of subtracted jets is stable with respect to the pileup which is a crucial condition for the jet reconstruction. Without satisfying this condition any cut applied on the jet  $p_T$  would lead to a choice of different subset of jets depending on the size of the pileup. Small deviation from zero of the jet momentum response is resulting from ignoring the rapidity dependence of the pileup density  $\rho$  and  $\rho_m$  and from a small average bias of pileup densities by a presence of the hard-scatter event. Such small deviation can be easily corrected after the jet reconstruction by multiplying the jet momentum by a correction factor. The constituent subtraction performs equally well as the area-based subtraction.

The small deviation of jet momentum response from zero for corrected jets can be contrasted with the jet momentum response of uncorrected jets — in the case of low pileup scenario the momentum response is 10-20%, in the case of high pileup scenario the jet momentum response is around 40%. This means that in the high pileup case, there is on average approximately 40 GeV of the pileup background underneath each jet leading to a reconstruction of a typical 100 GeV jet as a 140 GeV jet if performing no subtraction.

The second quantity characterizing the basic performance of the subtraction is the jet momentum resolution defined as  $\sigma[\Delta p_T] / p_T^{\text{orig}} = \sigma[p_T - p_T^{\text{orig}}] / p_T^{\text{orig}}$  where  $\sigma[\dots]$  denotes the standard deviation. The jet momentum resolution is dictated by the presence of fluctuations in the underlying pileup background leading to dependence  $\sigma[\Delta p_T] / p_T^{\text{orig}} = c / p_T^{\text{orig}}$  where  $c$  is a constant. The jet momentum resolution is shown in the upper middle panel of figure 1. The fit of the jet momentum resolution of uncorrected jets by  $c / p_T^{\text{orig}}$  leads to  $c \approx 25$  GeV which results from the magnitude of  $\sigma[\rho]$  and pileup fluctuations in  $\eta - \phi$  plane

---

<sup>3</sup>This pseudorapidity requirement limits the rate of top quarks originating from  $Z'$  boson decaying at rest. Thus, in this study, a significant fraction of top quarks originate from decays of  $Z'$  carrying a non-zero momentum in the direction of colliding beams.



**Figure 1:** Jet momentum response (left). Jet momentum resolution (upper middle) and jet finding efficiency (lower middle). Jet position resolution (right). Jets prior the subtraction denoted as pileup (square markers) are compared to jets corrected by the constituent subtraction (full circles) and jets corrected by the area-based subtraction (open circles). The dashed line in the left panel shows a constant at  $-0.8\%$  to guide the eye. Curve in the upper middle panel represents a fit of the jet momentum resolution by  $c/p_T$  resulting  $c \approx 25$  GeV.

within each particular event. The constituent subtraction and area-based subtraction have similar jet momentum resolution while both methods significantly improve it.

The jet finding efficiency is defined as the number of original jets having a matching corrected (or uncorrected) jet divided by the number of original jets. The matching criterion is the distance in the  $\eta - \phi$  plane between the original jet and corrected (or uncorrected) jet satisfying the condition  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.2$ . This quantity is plotted in the lower middle panel of figure 1. It shows that in the case of high pileup events it is difficult in principle to reconstruct the jets with  $p_T^{\text{orig}} < 50$  GeV due to the presence of sizable fluctuating background. The jet efficiency is better for the constituent subtraction than for the area-based subtraction.

The last quantity evaluated is the jet position resolution which characterizes the ability to recover the original jet axis in pseudorapidity,  $\eta$ , or azimuth,  $\phi$ . The jet position resolution in  $\eta$  and  $\phi$  are similar. The jet position resolution in  $\eta$ ,  $\sigma[\Delta\eta]$ , is defined as the standard deviation of the difference between the original jet  $\eta$  position and  $\eta$  position of the corrected (or uncorrected) jet. The  $\sigma[\Delta\eta]$  is plotted in the right panel of figure 1 where a clear difference between jets with pileup contribution and corrected jets is present. Jets corrected by the area-based subtraction have slightly worse jet position resolution than jets corrected by the constituent subtraction.

Based on the analysis of the basic performance, we can conclude the constituent subtraction has a good ability to correct for the pileup background and to recover the original jet kinematics even in the presence of a sizable pileup. The constituent subtraction has



generally similar performance as the area-based subtraction. A slightly better performance in terms of jet position resolution and jet efficiency may be attributed to the correction of the jet internal structure which is done by the constituent subtraction. The ability of the constituent subtraction to correct the jet internal structure is discussed in section 3.3.

### 3.2 Jet shape definitions

The ability of the constituent subtraction method to recover the internal structure of jet has been tested by evaluating four jet shape variables that are discussed in the literature to be useful for analyzing the boosted objects or to perform jet tagging [34, 35]. Here we briefly introduce these four jet shapes:

- The jet mass which can also be used to identify the hadronic decays of boosted heavy particles [38, 39].
- The  $N$ -subjettiness,  $\tau_N$ , defined as [36]

$$\tau_N = \frac{1}{d_0} \sum_k p_{T_k} \cdot \min(\Delta R_{1k}, \Delta R_{2k}, \dots, \Delta R_{Nk}), \quad \text{with,} \quad d_0 \equiv \sum_k p_{T_k} \cdot R \quad (3.2)$$

where  $R$  is the distance parameter of the jet algorithm,  $p_{T_k}$  is the transverse momentum of constituent  $k$  and  $\Delta R_{ik}$  is the distance between a subjet  $i$  and a constituent  $k$ . The  $N$  subjets are defined by re-clustering the constituents of the jet with exclusive version of the  $k_t$  algorithm [16] and requiring that exactly  $N$  subjets are found. Beside the  $N$ -subjettiness also the subjettiness ratio,  $\tau_{MN} = \tau_M/\tau_N$ , can be used to characterize the jet substructure. Typically, the three-to-two ratio,  $\tau_{32} = \tau_3/\tau_2$ , is used which provides a good discrimination between standard QCD jets and jets formed e.g. by boosted top quarks.

- The  $k_t$  splitting scale,  $\sqrt{d_{12}}$ , defined as [37]

$$\sqrt{d_{12}} = \min(p_T^1, p_T^2) \cdot \Delta R_{12}, \quad (3.3)$$

where  $p_T^1$  and  $p_T^2$  are the transverse momenta of two subjets and  $\Delta R_{12}$  is the distance between these two subjets. The two subjets are found by going back one step in the clustering history of the  $k_t$  algorithm. The variable  $\sqrt{d_{12}}$  can be used to distinguish heavy-particle decays, which tend to be reasonably symmetric when the decay is to like-mass particles, from the largely asymmetric splittings that originate from QCD radiation in light-quark or gluon jets.

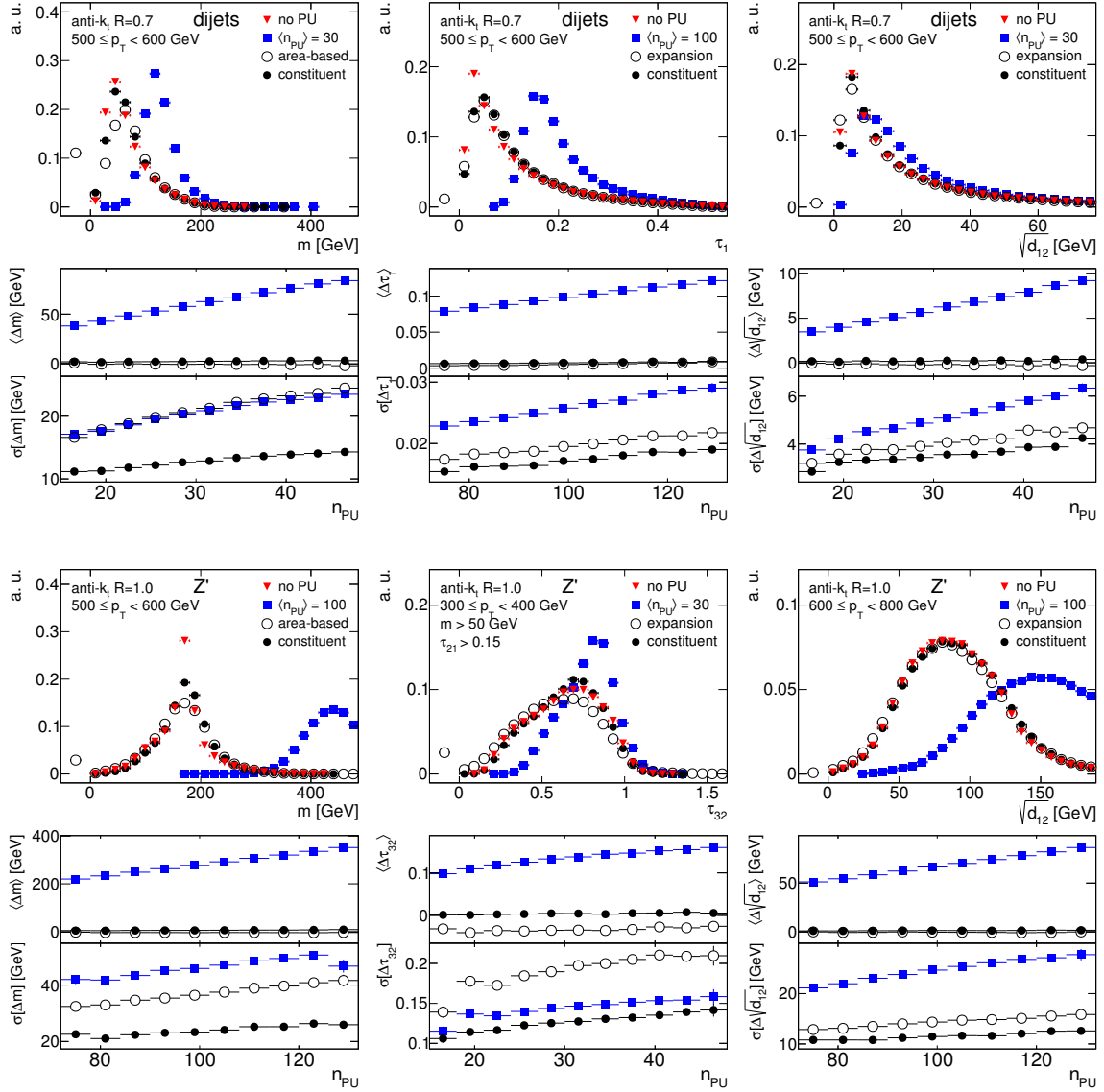
- The longitudinally invariant version of the planar flow, Pf, defined as [8]:

$$\text{Pf} = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2}, \quad (3.4)$$

where  $\lambda_1$  and  $\lambda_2$  are the eigenvalues of  $2 \times 2$  matrix:

$$M_{\alpha\beta} = \sum_i p_{T_i} \cdot (\alpha_i - \alpha_{\text{jet}}) \cdot (\beta_i - \beta_{\text{jet}}), \quad (3.5)$$

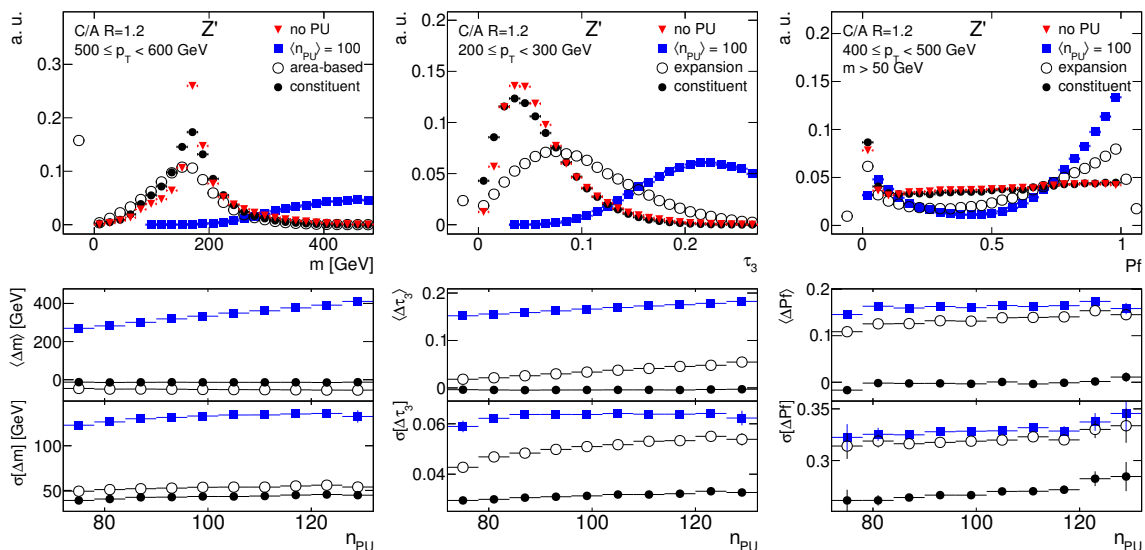
where  $\alpha$  and  $\beta$  correspond to rapidity or azimuth.



**Figure 2:** Performance of the constituent subtraction for jets clustered with the anti- $k_t$  algorithm. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and constituent subtraction method, respectively. The  $n_{PU}$  dependence of mean  $\langle \Delta x \rangle$  and standard deviation  $\sigma[\Delta x]$  are shown in the lower panel for each jet shape.

### 3.3 Jet shape subtraction

The constituent subtraction method is tested on various combinations of signal samples, pileup conditions, clustering algorithms, and jet shapes defined in section 3.2. The details of the configuration are provided at the beginning of section 3. The constituent subtraction can recover the original jet shape with a good accuracy in all evaluated combinations.



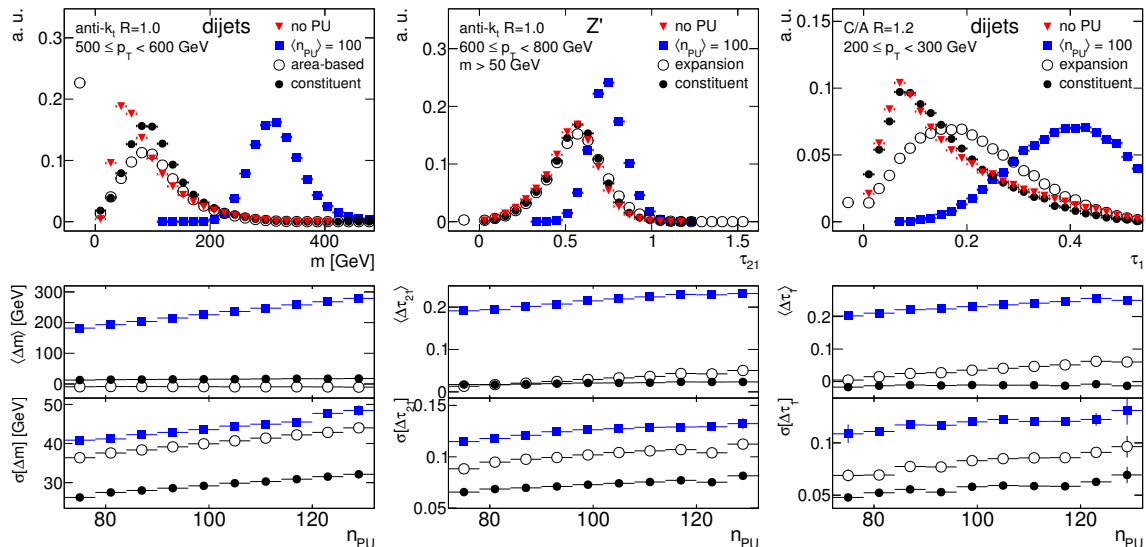
**Figure 3:** Performance of the constituent subtraction for jets clustered with the C/A algorithm. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and constituent subtraction method, respectively. The  $n_{PU}$  dependence of mean  $\langle \Delta x \rangle$  and standard deviation  $\sigma[\Delta x]$  are shown in the lower panel for each jet shape.

A representative subset of performance plots for the anti- $k_t$  algorithm is shown in figure 2 for various jet  $p_T$  intervals.<sup>4</sup> Four distributions of jet shapes are plotted for each presented plot: the original distribution (that is the distribution without pileup), the distribution with pileup, the distributions corrected by the constituent subtraction and the shape-expansion<sup>5</sup> methods. To quantify precisely the performance of the correction, two quantities have been evaluated for the differences between jet shape  $x$  and its original value without pileup  $x^{orig}$ : the mean value of these differences  $\langle \Delta x \rangle = \langle x - x^{orig} \rangle$  and the standard deviation of these differences  $\sigma[\Delta x] = \sigma[x - x^{orig}]$  which represents the resolution. For each combination of configurations, the uncorrected distributions differ significantly from the corresponding original distribution, and have a significant dependence on  $n_{PU}$ . A substantial improvement is achieved by the constituent subtraction. The mean difference  $\langle \Delta x \rangle$  does not exhibit the  $n_{PU}$  dependence and it is always centered near zero after the subtraction. The resolution  $\sigma[\Delta x]$  is improved as well. The constituent subtraction method performs similarly or mostly better when compared to the shape-expansion method.

For any of the studied jet shapes, the shape-expansion method can lead to negative corrected jet shapes that are unphysical. To better visualize the contribution of such

<sup>4</sup>The  $p_T$  of the jets without pileup is used to define the  $p_T$  intervals. The jets with pileup and the corrected jets are matched to the original jets without pileup. In some cases, additional cut is applied on mass or subjettiness ratio  $\tau_{21}$  again on jets without pileup.

<sup>5</sup>For the jet mass, the area-based method using eq. (3.1) is used which is identical to the shape-expansion method.



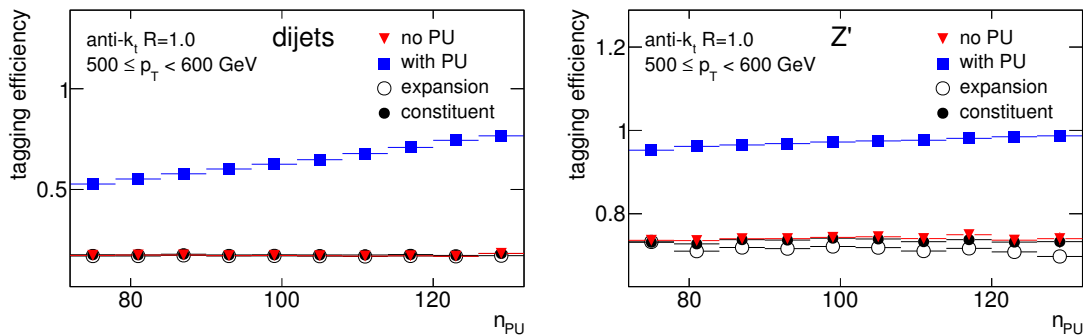
**Figure 4:** Performance of the constituent subtraction in events simulating a segmented detector. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and constituent subtraction method, respectively. The  $n_{PU}$  dependence of mean  $\langle \Delta x \rangle$  and standard deviation  $\sigma[\Delta x]$  are shown in the lower panel for each jet shape.

values, the first bin with negative jet shape in plots of figures 2–4 is set to the fraction of negatively corrected jet shapes.<sup>6</sup> Unphysical values can also occur in the case of the area-based correction of the jet mass when the corrected energy is smaller than the magnitude of corrected momentum. Again, the negative bin represents the fraction of such jets. The fraction of unphysical jet shapes obtained from the shape-expansion method reaches up to  $\sim 12\%$  depending on the  $p_T$  interval and the type of the jet shape.

The constituent subtraction method has been tested also on the jets clustered with C/A algorithm which is often employed in various studies of the jet substructure and boosted objects [34]. The clustering in the C/A algorithm is based purely on the geometry and thus it leads to jet with a different jet area compared to the anti- $k_t$  algorithm [12]. The performance of the constituent subtraction method for C/A algorithm with distance parameter  $R = 1.2$  is shown in figure 3. For this configuration, the impact of the pileup on jet shapes is much stronger compared to the configuration with the anti- $k_t$  algorithm. The constituent subtraction can recover the original distributions and it exhibits significantly better ability to subtract the pileup compared to the shape-expansion method.

Further, the constituent subtraction method has been tested on jets reconstructed in events run through a simple simulation of a segmented detector. In this simulation, the  $\eta - \phi$  plane is divided into cells of size  $0.1 \times 0.1$ . Particles pointing to the same cell are

<sup>6</sup>For the shape-expansion correction of ratios  $\tau_{21}$  or  $\tau_{32}$ , the numerator and denominator are corrected individually. When at least one of these corrected variables is negative, the corrected  $\tau_{21}$  or  $\tau_{32}$  is counted as negative. For the calculation of the mean and resolution of  $\tau_{21}$  or  $\tau_{32}$ , the negative values are not used. For any other jet shapes, the negative values are set to zero so that they do not bias the mean and resolution.



**Figure 5:** Tagging efficiencies for background (left) and signal (right) samples using tagging cut  $\sqrt{d_{12}} > 50$  GeV in events simulating a segmented detector. Red triangles show the tagging efficiency for jets without pileup. Blue squares show the tagging efficiency for jets with pileup, open and closed circles show tagging efficiency for jets corrected by the shape-expansion method and constituent subtraction method, respectively.

combined into one new effective particle by summing their energies. The mass of the cell is set to zero and its  $\eta - \phi$  position is set to the center of the cell. These new effective particles have the same properties as the calorimeter clusters or towers used in real experiments and they are a combination of the pileup and signal. The jet finding algorithm runs over these events delivering jets that are corrected in the same way as for events composed of standard particles. A typical example of the performance of the subtraction methods in case of this simulation is shown in figure 4. The constituent subtraction exhibits very similar performance as without the detector simulation which also applies for the shape-expansion method while the constituent subtraction again outperforms the shape-expansion method. Degradation of the performance of jet mass reconstruction is seen for all methods in the configuration of jet reconstruction with large distance parameter. This is a consequence of a sensitivity of jet mass to variations in the constituent transverse momenta of constituents that are at large angles with respect to the jet axis. While other variables are less sensitive to such variations, the jet mass can change significantly by adding a soft particle to the jet periphery.

An important test is to evaluate the jet tagging performance. The splitting scale,  $\sqrt{d_{12}}$ , is used to tag the boosted top quarks from the  $Z' \rightarrow t\bar{t}$  decay and tagging efficiencies have been evaluated. The signal sample can be compared with the di-jet sample which in this case provides a reasonable estimate of the background for the  $Z' \rightarrow t\bar{t}$  decay. The result is shown on figure 5. One can see that both the constituent subtraction and shape-expansion can achieve the same tagging efficiency as in the case of no pileup.

The above presented results demonstrate the stability and good performance of the constituent subtraction method.

#### 4 Conclusions

We have introduced a new tool to correct for the pileup in high-luminosity LHC running that represents an extension and a simplification of the current state of the art. The

constituent subtraction method operates at the level of the jet constituents and provides both a performance improvement and a simplification compared to existing methods: the precision of the reconstruction of jet shapes is improved as well as the speed of the correction itself.

The constituent subtraction method is tested by evaluating the pileup dependence, and other key metrics, of several jet shapes and jet kinematics using multiple jet definitions. Improvements are demonstrated in both the reduction of fluctuations in the resulting jet shapes as a function of pileup and the ability to remove the pileup dependence of the corrected quantities. It also provides a better jet position resolution and jet finding efficiency for reconstructing jets from boosted objects, which directly impacts the experimental sensitivity to new physics.

Since the correction proceeds without knowledge of a jet algorithm, a novel application of this approach would be to correct the whole event prior to jet finding. Not only would this have implications for the technical performance and computational resources currently devoted to evaluation the pileup subtraction for each jet individually, but it also has the potential to improve the determination of the missing transverse energy. Furthermore, the constituent subtraction approach can be used to correct for the underlying event in heavy ion collisions.

It is very important that these studies be verified by the experiments using both fully simulated data samples of signal and background events, as well as *in situ* studies using data. Given the excellent correspondence between the similar preliminary studies of earlier pileup correction methods and the experimental reality, we expect that the performance observed here is quantitatively representative of what can be achieved by the experimental collaborations. Nonetheless, any new approach must be vetted and tested thoroughly.

## Acknowledgments

The authors would like to thank Gregory Soyez and Gavin Salam for their feedback and assistance in understanding and using the shape-expansion tools. The work is supported by Charles University in Prague, projects PRVOUK P45, UNCE 204020/2012, and GA UK No 320813. DWM is supported in part by the Neubauer Family Foundation Fellows Program for Assistant Professors.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] ATLAS collaboration, *Pile-up corrections for jets from proton-proton collisions at  $\sqrt{s} = 7$  TeV in ATLAS in 2011*, [ATLAS-CONF-2012-064](#) (2012).
- [2] J.M. Butterworth, A.R. Davison, M. Rubin and G.P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [[arXiv:0802.2470](#)] [[INSPIRE](#)].

- [3] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev. D* **80** (2009) 051501 [[arXiv:0903.5081](#)] [[INSPIRE](#)].
- [4] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [[arXiv:0912.1342](#)] [[INSPIRE](#)].
- [5] D. Krohn, M.D. Schwartz, M. Low and L.-T. Wang, *Jet Cleansing: Pileup Removal at High Luminosity*, [arXiv:1309.4777](#) [[INSPIRE](#)].
- [6] CMS collaboration, *Jet performance in CMS*, CERN-CMS-CR-2013-325 (2013).
- [7] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* **659** (2008) 119 [[arXiv:0707.1378](#)] [[INSPIRE](#)].
- [8] G. Soyez, G.P. Salam, J. Kim, S. Dutta and M. Cacciari, *Pileup subtraction for jet shapes*, *Phys. Rev. Lett.* **110** (2013) 162001 [[arXiv:1211.2811](#)] [[INSPIRE](#)].
- [9] L. Apolinario, N. Armesto and L. Cunqueiro, *An analysis of the influence of background subtraction and quenching on jet observables in heavy-ion collisions*, *JHEP* **02** (2013) 022 [[arXiv:1211.1161](#)] [[INSPIRE](#)].
- [10] M. Spousta, *Jet Quenching at LHC*, *Mod. Phys. Lett. A* **28** (2013) 1330017.
- [11] CMS collaboration, *Underlying-Event Subtraction for Particle Flow*, CMS DP-2013/018.
- [12] M. Cacciari, G.P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005 [[arXiv:0802.1188](#)] [[INSPIRE](#)].
- [13] G. Bellettini, R. Bertani, C. Bradaschia, R. Del Fabbro, A. Scribano et al., *Hadron calorimeter towers with a high space resolution*, *Nucl. Instrum. Meth.* **204** (1982) 73 [[INSPIRE](#)].
- [14] ATLAS collaboration, *Calorimeter Clustering Algorithms : Description and Performance*, ATL-LARG-PUB-2008-002 (2008).
- [15] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [[hep-ph/9305266](#)] [[INSPIRE](#)].
- [16] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187 [[INSPIRE](#)].
- [17] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [18] G.C. Blazey, J.R. Dittmann, S.D. Ellis, V.D. Elvira, K. Frame et al., *Run II jet physics*, FERMILAB-CONF-00-092-E [[hep-ex/0005012](#)] [[INSPIRE](#)].
- [19] D.E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, *Phys. Rev. D* **87** (2013) 054012 [[arXiv:1211.3140](#)] [[INSPIRE](#)].
- [20] <http://fastjet.hepforge.org/contrib>.
- [21] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](#)] [[INSPIRE](#)].
- [22] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [23] CTEQ collaboration, H.L. Lai et al., *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, *Eur. Phys. J. C* **12** (2000) 375 [[hep-ph/9903282](#)] [[INSPIRE](#)].

- [24] M. Cacciari, G.P. Salam and G. Soyez, *The Anti- $k(t)$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [25] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [[hep-ph/9707323](#)] [[INSPIRE](#)].
- [26] 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** (2013) 076 [[arXiv:1306.4945](#)] [[INSPIRE](#)].
- [27] CMS Collaboration, *Jet Performance in  $pp$  Collisions at 7 TeV*, CMS-PAS-JME-10-003 (2010).
- [28] O.S. Bruning, *HL-LHC Parameter Space and Scenarios*, *Proc. Chamonix 2012 workshop on LHC Performance* (2012).
- [29] M. Cacciari and G.P. Salam, *Dispelling the  $N^3$  myth for the  $k_t$  jet-finder*, *Phys. Lett. B* **641** (2006) 57 [[hep-ph/0512210](#)] [[INSPIRE](#)].
- [30] ATLAS collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Eur. Phys. J. C* **73** (2013) 2304 [[arXiv:1112.6426](#)] [[INSPIRE](#)].
- [31] CMS collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, *2011 JINST* **6** P11002.
- [32] CDF collaboration, A. Bhatti et al., *Determination of the jet energy scale at the collider detector at Fermilab*, *Nucl. Instrum. Meth. A* **566** (2006) 375 [[hep-ex/0510047](#)] [[INSPIRE](#)].
- [33] D0 collaboration, *Jet energy scale determination in the D0 experiment*, *FERMILAB-PUB-13-577* (2013).
- [34] A. Altheimer et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys. G* **39** (2012) 063011.
- [35] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [[arXiv:1106.3076](#)] [[INSPIRE](#)].
- [36] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with  $N$ -subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [37] J.M. Butterworth, B.E. Cox and J.R. Forshaw,  *$WW$  scattering at the CERN LHC*, *Phys. Rev. D* **65** (2002) 096014 [[hep-ph/0201098](#)] [[INSPIRE](#)].
- [38] A. Abdesselam, E.B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J. C* **71** (2011) 1661 [[arXiv:1012.5412](#)] [[INSPIRE](#)].
- [39] CMS collaboration, *Studies of jet mass in dijet and  $W/Z$ +jet events*, *JHEP* **05** (2013) 090 [[arXiv:1303.4811](#)] [[INSPIRE](#)].