

RECEIVED: January 3, 2013

REVISED: March 5, 2013

ACCEPTED: March 9, 2013

PUBLISHED: March 28, 2013

## Jet substructure by accident

---

Timothy Cohen,<sup>a</sup> Eder Izaguirre,<sup>b</sup> Mariangela Lisanti<sup>c</sup> and Hou Keong Lou<sup>d</sup>

<sup>a</sup>*SLAC, Stanford University,  
Menlo Park, CA 94025, U.S.A.*

<sup>b</sup>*Perimeter Institute for Theoretical Physics,  
Waterloo, ON, N2L 2Y5, Canada*

<sup>c</sup>*PCTS, Princeton University,  
Princeton, NJ 08544, U.S.A.*

<sup>d</sup>*Department of Physics, Princeton University,  
Princeton, NJ 08540, U.S.A.*

*E-mail:* [timcohen@slac.stanford.edu](mailto:timcohen@slac.stanford.edu), [eizaguirre@perimeterinstitute.ca](mailto:eizaguirre@perimeterinstitute.ca),  
[mlisanti@princeton.edu](mailto:mlisanti@princeton.edu), [hlou@princeton.edu](mailto:hlou@princeton.edu)

**ABSTRACT:** We propose a new search strategy for high-multiplicity hadronic final states. When new particles are produced at threshold, the distribution of their decay products is approximately isotropic. If there are many partons in the final state, it is likely that several will be clustered into the same large-radius jet. The resulting jet exhibits substructure, even though the parent states are not boosted. This “accidental” substructure is a powerful discriminant against background because it is more pronounced for high-multiplicity signals than for QCD multijets. We demonstrate how to take advantage of accidental substructure to reduce backgrounds without relying on the presence of missing energy. As an example, we present the expected limits for several  $R$ -parity violating gluino decay topologies. This approach allows for the determination of QCD backgrounds using data-driven methods, which is crucial for the feasibility of any search that targets signatures with many jets and suppressed missing energy.

**KEYWORDS:** Jets, Supersymmetry Phenomenology

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Quantifying accidental substructure</b>	<b>5</b>
2.1	Jet mass	5
2.2	$N$ -subjettiness	6
2.3	Introducing event-subjettiness	7
<b>3</b>	<b>Analysis strategy</b>	<b>8</b>
3.1	Event generation	9
3.2	Expected reach	10
<b>4</b>	<b>Conclusion</b>	<b>14</b>
<b>A</b>	<b>Simulation details and validation</b>	<b>16</b>

---

**1 Introduction**

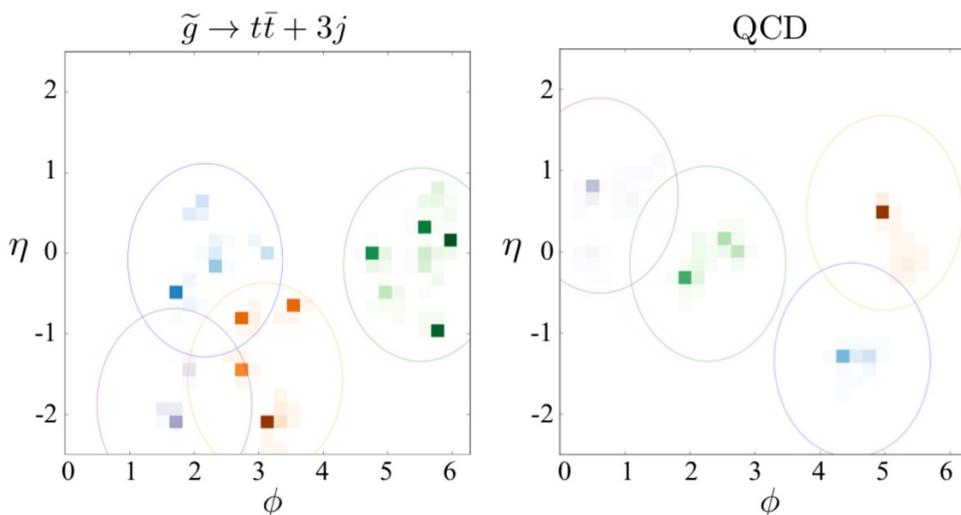
Our approach to jet physics is undergoing a renaissance. While most LHC studies use the energy and momentum of a jet, there is growing appreciation for the wealth of information that can be extracted by analyzing a jet’s internal structure (see [1–3] for reviews). Jet substructure gained traction when it was shown to increase the LHC sensitivity to Higgs boson decays into  $b$ -quarks [4]. Since then, jet substructure has been applied by theorists to a variety of scenarios [5–27], and its power has been demonstrated experimentally in Tevatron [28, 29] and LHC [30–36] searches.

In all existing studies, jet substructure has been used to search for boosted resonances with collimated decay products that are reconstructed as a single jet. For a typical event at the LHC, parent particles are produced near threshold; the decay products are boosted for the small fraction of signal events produced with significant transverse momentum,<sup>1</sup> or in the case where the parent particle decays to significantly lighter daughters. In this paper, we explore a new application for jet substructure techniques that does not rely on having collimated decay products.

We demonstrate that substructure technology is useful in the non-boosted regime for models that yield a high multiplicity of hadronic final states. This strategy relies on the fact that when new particles with  $\mathcal{O}(\text{TeV})$  masses are produced at threshold, their decay products tend to be distributed isotropically in the detector. Our proposal requires an event to contain several (specifically, four or more) large-radius jets defined using the anti- $k_T$  algorithm [39] with angular size  $R = 1.2$ . Because these so-called “fat” jets can cover a

---

<sup>1</sup>For example, the signal efficiency when targeting boosted gluinos is roughly  $\mathcal{O}(\text{few } \%)$  at the LHC [37, 38].



**Figure 1.** Lego plots showing the distribution of calorimeter activity in the  $\eta - \phi$  plane. The different colors correspond to different fat jets; within each panel, darker colors signify higher  $p_T$  in a given detector cell. Note that the relative  $p_T$  scale is different for the signal and background example. The signal (left panel) is pair production of 500 GeV gluinos with  $\tilde{g} \rightarrow t\bar{t} + 3j$ , which yields up to 18 partons in the final state. The gluinos have transverse momenta of 120 and 65 GeV, so they are essentially at rest. A QCD multijet event is depicted in the right panel. The circles are centered on the clustered fat jet with a radius of  $R = 1.2$  to schematically illustrate the extent of each fat jet. There is significant substructure for the signal and suppressed substructure for the background.

large fraction of the effective detector area, several decay partons from a high-multiplicity signal will often get clustered into a single fat jet. Non-boosted final states can therefore manifest “accidental substructure.”

Requiring multiple fat jets with non-trivial substructure greatly reduces QCD contamination. For an event to have four fat jets, it must have at least this many well-separated hard partons. The presence of substructure in the remaining QCD sample is most likely to occur when one or more isolated partons undergoes a hard  $1 \rightarrow 2$  splitting. Because this process is dominated by a soft and/or collinear singularity, the probability decreases as the energy and separation of the final states increases. As a result, QCD events typically have suppressed substructure.

Figure 1 illustrates why accidental substructure is useful for distinguishing between a typical signal and background event. These “lego plots” show the spatial distribution of calorimeter activity in the  $\eta - \phi$  plane, where  $\eta$  is pseudorapidity and  $\phi$  is azimuthal angle. The left panel is a lego plot for a signal event with up to 18 partons in the final state; the signal is gluino pair production, where each gluino decays to a pair of top quarks and an unstable neutralino that decays to three partons (see the left diagram in figure 2). The right panel shows the lego plot for a QCD event. The different colors correspond to different fat jets in the event. It is clear that the fat jets from signal have more pronounced substructure than the ones from QCD.

Figure 1 suggests that cutting on the number of small-radius ( $R \sim 0.4$ ) jets may suf-

face to distinguish signal from background. An explicit high jet multiplicity search requires accurate modeling of the QCD background, which has intrinsic theoretical challenges. The current state of the art is tree-level QCD calculations that rely on matrix element-parton shower matching up to six jets. Because additional jets must be generated by the parton shower, these calculations systematically underestimate the  $p_T$  spectrum of the high multiplicity tail. Higher multiplicity, matched, next-to-leading order calculations are not anticipated in the near future, implying that precision modifications to the shapes of the QCD distribution will not be known. Finally, even once this has been achieved, there is the computational limitation associated with populating the entire  $3-n$  dimensional phase space for events with  $n$  jets. As a result, theorists should validate Monte Carlo background predictions against data to derive plausible limits. There exist studies from the CMS and ATLAS collaborations that present 6 jet [37, 40] and 8 jet [41] distributions. However, these do not provide enough information to place cuts on the number of small-radius jets larger than  $\sim 6-8$ . This constrains theoretical investigations of high multiplicity searches with small-radius jets.

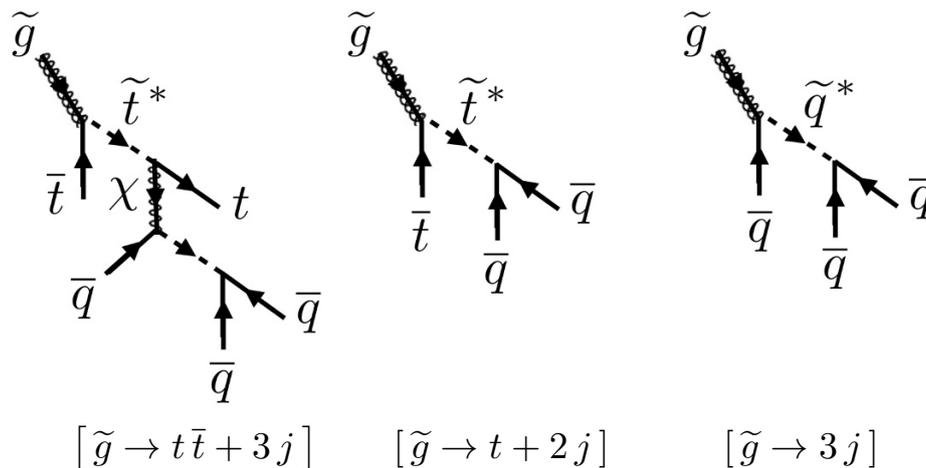
An experimental analysis targeting many small-radius jets must obtain the multijet backgrounds from data. Current data-driven methods for determining detailed kinematic features of small-radius jets are limited in that they rely on ad hoc fitting functions to perform background extrapolations. If a search that utilized these procedures yields an excess of events, there is no guidance for investigating the discrepancy because the functions are not derived from an underlying theory.<sup>2</sup>

Searches that use fat jets can implement an alternate strategy to estimate backgrounds. For the substructure analysis proposed here, one can study the internal structure of fat dijets. Because this sample should be signal poor, it can be used to determine the pure QCD dependence of jet mass and substructure on other quantities like jet  $p_T$ . These results can then be extrapolated to four fat jet events, and should lead to reasonable background predictions so long as the correlations between fat jets are small. Importantly, the associated systematics for a fat jet analysis differ from those that dominate in a search for many small-radius jets. It is beneficial to have competing searches with different systematics to ensure that new physics is not overwhelmed by large uncertainties.

Finally, we note that our analysis does not rely on the presence of missing transverse energy ( $\cancel{E}_T$ ), which is typically crucial for discriminating against multijet backgrounds in searches for supersymmetry (SUSY). Missing energy is not a robust prediction of SUSY models, e.g.  $R$ -parity can be violated, the superpartner spectrum can be squeezed, or SUSY can be stealthy [44, 45]. There are also a number of non-SUSY models that have signatures without  $\cancel{E}_T$ , such as [46–49]. To cover these and other  $\cancel{E}_T$ -less theories at the LHC, it is imperative to develop new search strategies to efficiently reduce the QCD background. Such a strategy could rely on rare objects in the event, such as  $b$ -jets or leptons, to further reduce backgrounds. However, a search that is independent of these extra handles is powerful for its generality. Because our proposal *only* relies on having a final state with many jets, it can be used to place limits on a wide-range of model space.

---

<sup>2</sup>For recent theoretical progress on extrapolating jet multiplicity, see [42, 43].



**Figure 2.** Gluino decay diagrams, illustrating topologies that can lead to as many as 18, 10, and 6-parton final states (left to right, respectively) when the gluinos are pair-produced. Note that  $\tilde{g}$  is a gluino,  $\tilde{t}$  is a stop,  $t$  is a top quark,  $\tilde{q}$  is a first or second generation squark,  $\chi$  is a neutralino, and  $j$  refers to a final state quark or anti-quark.

We demonstrate that accidental substructure is a powerful discriminator by applying it to three distinct gluino  $\tilde{g}$  decay scenarios when the  $R$ -parity violating (RPV) superpotential coupling  $U^c D^c D^c$  is non-zero:

$$\tilde{g} \rightarrow t\bar{t} + 3j, \quad \tilde{g} \rightarrow t + 2j, \quad \text{and} \quad \tilde{g} \rightarrow 3j. \tag{1.1}$$

Here  $j$  refers to a final state quark or anti-quark, not to a detector-level jet. When the gluinos are pair-produced, these three topologies can lead to as many as 18, 10, and 6-parton final states, respectively, as shown in figure 2. The first topology arises when a gluino decays to a pair of tops and an unstable neutralino, which decays to three partons through an off-shell squark via  $U^c D^c D^c$ . The other two topologies correspond to the RPV gluino decays into  $tbs$  and  $uds$  final states. For a review of constraints on these RPV interactions, see [50]. The 18 and 10-parton topologies are particularly well-motivated theoretically because the top quarks in the final state can result from a light stop in the spectrum. This is a plausible scenario with minimal fine-tuning where the non-zero RPV couplings suppress  $\cancel{E}_T$ , thereby hiding SUSY from current searches [51]. In particular, the 10-parton topology was the focus of a recent proposal that used substructure techniques to look for boosted stops [26].

The remainder of this paper proceeds as follows. In section 2, we present the needed variables, jet mass and  $N$ -subjettiness, and introduce the concept of “event-subjettiness.” In section 3, we show how these tools can be combined into a full analysis. After a brief description of the event generation procedure, we present the expected limits for the different gluino decay topologies. We conclude in section 4. Appendix A contains a detailed description of our simulations, including validation plots.

## 2 Quantifying accidental substructure

Our analysis relies on two observables: total jet mass and event-subjettiness. The latter is a new variable that we introduce to quantify the amount of accidental substructure in an event. It requires  $N$ -subjettiness to characterize the subjet nature of each jet. Jet mass,  $N$ -subjettiness, and event-subjettiness form the cornerstone of our analysis, so we introduce them individually here. The full analysis strategy is presented in section 3 and the details of our Monte Carlo event generation, detector mock-up, and validation can be found in section 3.1 and appendix A.

For the figures in this section, we select 8 TeV LHC events with at least four jets, clustered using the anti- $k_T$  algorithm [39] with cone size  $R = 1.2$ . The transverse momenta of the leading and subleading fat jets must satisfy  $p_T \geq 100$  GeV and  $p_T \geq 50$  GeV, respectively. Although no 8 TeV multijet,  $\cancel{E}_T$ -less triggers are publicly available, some 7 TeV examples include: five or more jets ( $R = 0.4$ ) with  $p_T > 30$  GeV at ATLAS [37],  $\sim 500$ –750 GeV of  $H_T$  at CMS [41], and 4, 6, or 8 high- $p_T$  jets ( $R = 0.5$ ) at CMS [52]. We have verified that the first of these triggers is 100% efficient for the QCD background and the gluino topologies we consider after final selection cuts.

### 2.1 Jet mass

Standard SUSY searches at ATLAS and CMS use a combination of missing energy,  $\cancel{E}_T$ , and visible transverse energy,

$$H_T = \sum_{j=1}^{N_j} \sqrt{(p_T^2)_j + m_j^2}, \quad (2.1)$$

where  $j$  is a jet in the event with mass  $m_j \equiv \sqrt{E_j^2 - |\vec{p}_j|^2}$  and  $N_j$  is the number of jets in the event with  $p_T > 50$  GeV. The total jet mass of an event,

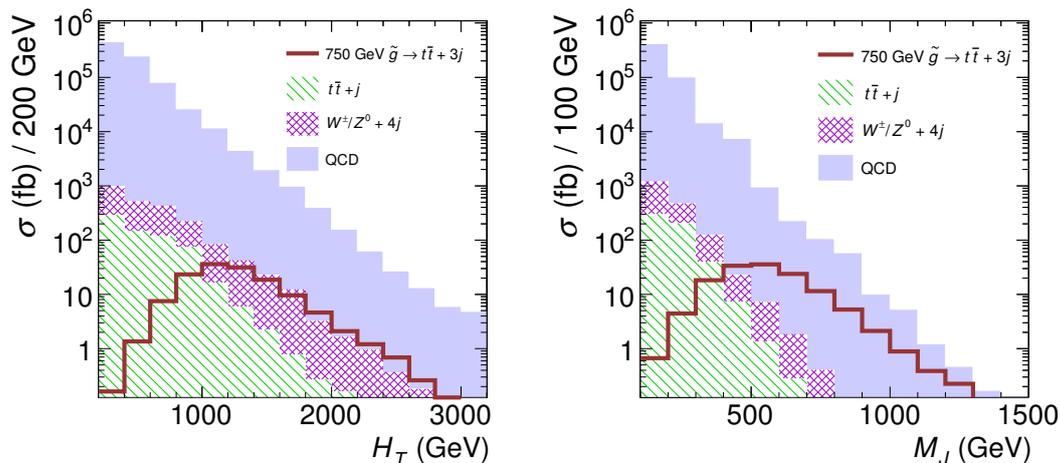
$$M_J \equiv \sum_{j=1}^{N_j} m_j, \quad (2.2)$$

is a more powerful discriminator than  $H_T$  in searches for high multiplicity final states [53] because a jet's mass automatically encodes gross kinematic features of its constituents.

Consider a small-radius jet that is seeded from an isolated parton. In the absence of showering, this jet will have zero mass. Non-zero jet mass arises if multiple partons are clustered together and/or from QCD radiation — the former yields a larger jet mass than the latter. As a result, a QCD and signal event with equivalent  $H_T$  can have different total jet mass. More quantitatively,  $H_T$  can be related to  $M_J$  via

$$H_T = \sum_{j=1}^{N_j} \sqrt{(p_T^2)_j + m_j^2} \propto \sum_{j=1}^{N_j} \sqrt{\langle m_j^2 \rangle ((\kappa R)^{-2} + 1)} \simeq M_J \frac{\sqrt{1 + (\kappa R)^2}}{\kappa R}, \quad (2.3)$$

where  $\kappa \simeq \sqrt{\alpha_s}$  for jets whose mass is generated from the parton shower [1] and  $\kappa \simeq 1$  for fat jets that contain multiple hard partons accidentally clustered in the same jet. Figure 3



**Figure 3.** The  $H_T$  (left) and  $M_J$  (right) distributions for the backgrounds and an example signal. The signal (red solid line) is pair production of a 750 GeV gluino with  $\tilde{g} \rightarrow t\bar{t} + 3j$ . The stacked histogram is for background (QCD in solid blue,  $W^\pm/Z^0 + 4j$  in hatched magenta, and  $t\bar{t} + j$  in striped green).  $M_J$  is a more powerful discriminator than  $H_T$  when comparing signal to background.

shows the  $H_T$  and  $M_J$  distributions for background and a signal example. Clearly, a cut on  $M_J$  improves sensitivity to the signal as opposed to an  $H_T$  requirement.

The authors of [53] proposed a study that took advantage of total jet mass for high multiplicity signals, but which still relied on a missing energy cut. In this work, we demonstrate that accidental substructure increases sensitivity when used in conjunction with total jet mass. This result is especially useful in topologies with  $\cancel{E}_T$  suppression, such as the benchmarks presented in figure 2. Adding a moderate  $\cancel{E}_T$  cut for other topologies that do contain sources of missing energy, e.g. new physics signals with tops in the final state, can provide an additional handle for improving the discriminating power of accidental substructure and jet mass.

## 2.2 $N$ -subjettiness

To quantify accidental substructure, we begin by considering the  $N$ -subjettiness variable  $\tau_N$  [14, 19, 54].  $\tau_N$  is a measure of the degree to which a fat jet has  $N$  well-separated subjects. For each jet,  $\tau_N$  is defined as

$$\tau_N = \frac{1}{d_\beta} \sum_i (p_T)_i \min \left\{ \Delta R_{i,1}^\beta, \Delta R_{i,2}^\beta, \dots, \Delta R_{i,N}^\beta \right\}$$

$$d_\beta = \sum_i (p_T)_i R_0^\beta, \tag{2.4}$$

where the minimization is performed by varying  $N$  axes,  $R_0$  is the choice of clustering radius, and  $\Delta R_{i,M} = \sqrt{(\Delta\phi_{i,M})^2 + (\Delta\eta_{i,M})^2}$  denotes the angular distance between the  $i^{\text{th}}$  constituent particle and the  $M^{\text{th}}$  axis. We take  $\beta = 1$  and  $R_0 = 1.2$ .

To elucidate what  $N$ -subjettiness measures, consider  $\tau_3$ . If the jet consists of three or fewer well-collimated subjects,  $\tau_3 \simeq 0$  because  $\min\{\Delta R_{i,1}, \Delta R_{i,2}, \Delta R_{i,3}\}$  vanishes for the  $i^{\text{th}}$

constituent. If the fat jet contains more than three subjets (or the particles making up the jet are not well-collimated),  $\tau_3 > 0$  because at least one subjet is not aligned with an axis.

While the individual  $\tau_N$  are not typically useful, ratios are [14]. For example,

$$\tau_{NM} \equiv \tau_N / \tau_M \tag{2.5}$$

is efficient at selecting  $N$ -subjettiness events for  $M < N$ . For a jet with  $N$  well-separated subjets,  $\tau_M$  is large,  $\tau_N$  is small, and therefore  $\tau_{NM}$  is much less than 1. Rejecting events with  $\tau_{NM} \sim 1$  selects for jets that are more  $N$ -prong like.<sup>3</sup>

Figure 4 shows the normalized distributions of  $\tau_{43}$  for each of the four hardest jets for QCD and the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology. The jets in each event are ordered by decreasing  $p_T$ . The background sample is peaked around  $\tau_{43} \sim 0.7$ – $0.8$ . In contrast, the distribution for the signal is shifted to lower values, with a tail that extends to lower  $\tau_{43}$ . These distributions reinforce the general conclusions we drew from the lego plots in figure 1. Specifically,  $\tau_{43}$  is shifted towards lower values for the signal relative to the background, suggesting that signal jets typically look more four-subjettiness than the background jets.

### 2.3 Introducing event-subjettiness

$N$ -subjettiness is useful for characterizing the number of subjets in a single jet. However, it would be useful to have a variable that takes into account the relative abundance of jets with substructure in an entire event. To this end, we introduce “event-subjettiness,”  $T_{NM}$ , which is defined as the geometric mean of the  $\tau_{NM}$  for the four hardest jets in an event:

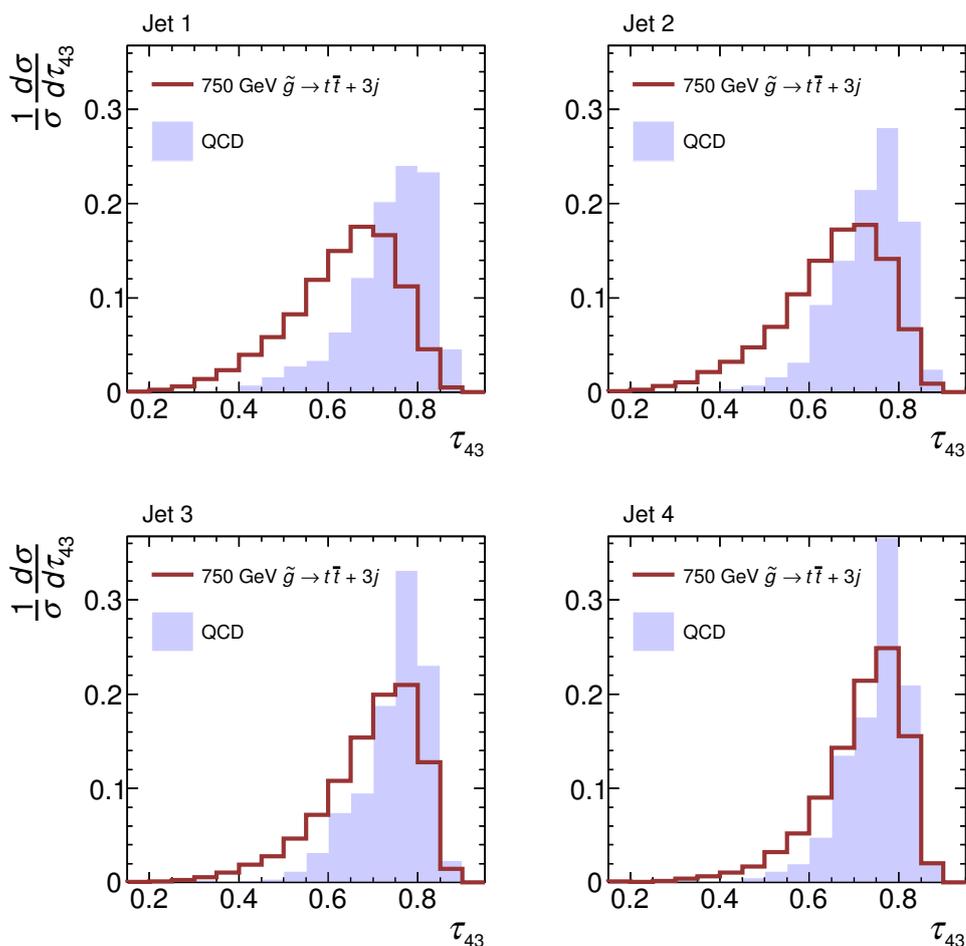
$$T_{NM} = \left[ \prod_{j=1}^4 (\tau_{NM})_j \right]^{1/4} . \tag{2.6}$$

The more jets with substructure in an event, the more jets with a small  $\tau_{NM}$ , resulting in a smaller value of  $T_{NM}$ . The geometric mean is less sensitive to the presence of a single high  $\tau_{NM}$  in an event than the arithmetic mean. In particular, the arithmetic (geometric) mean tends to result in slightly larger  $S/B$  ( $S/\sqrt{B}$ ) than the geometric (arithmetic) mean. This leads to a mild improvement in the reach when using the geometric mean. We also explored placing cuts on combinations of the  $\tau_{NM}$  for the single two hardest jets; this does not lead to the same level of discriminating power because the amount of substructure is not necessarily correlated with the hardness of a jet. As discussed in section 2.1, two jets with equivalent  $p_T$  can have a different mass depending on whether the mass arises from accidental substructure or hard-emission.

Figure 5 illustrates the distributions of  $T_{43}$  for backgrounds and the signal example with  $\tilde{g} \rightarrow t\bar{t} + 3j$ . For this topology, many of the signal fat jets often have four or more subjets, which drives down  $T_{43}$  relative to that for the backgrounds. This is evident, for

---

<sup>3</sup>Ensuring that the  $\tau_{NM}$  variables are infrared and collinear (IRC) safe [55] is important for the implementation of our proposed search.  $\tau_{N(N-1)}$  is IRC safe if applied to a jet with  $N$  hard prongs, which is ensured by requiring a lower bound on  $\tau_{(N-1)(N-2)}$ . While the results presented here do not include this requirement, we have verified that they are not significantly changed by a naive application of the IRC safety conditions.

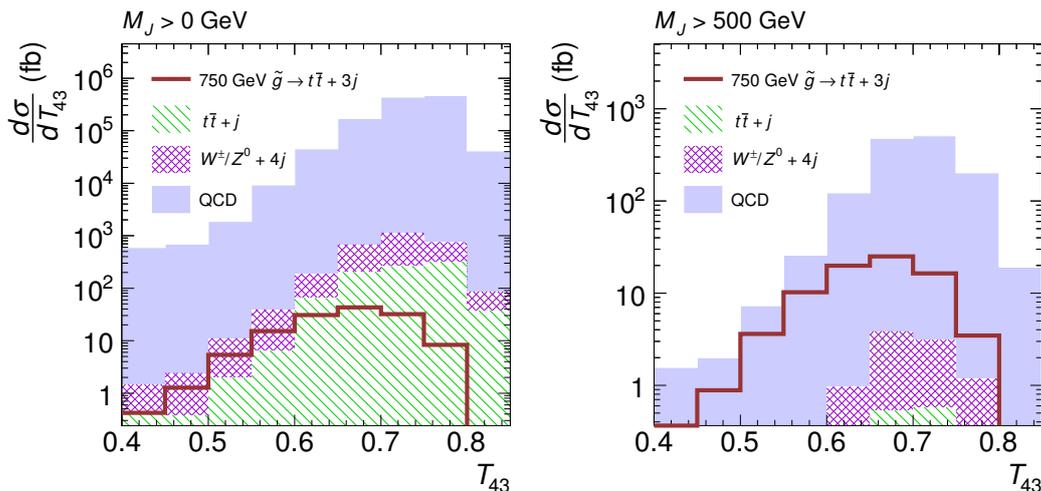


**Figure 4.** Normalized distributions of  $\tau_{43}$  for background and a signal example. Each plot shows the normalized distribution before a cut on  $M_J$ . The signal (red solid line) is pair production of a 750 GeV gluino with  $\tilde{g} \rightarrow t\bar{t} + 3j$ . The solid blue histogram is for the QCD background. Each panel is the distribution for the  $j^{\text{th}}$  jet; the order is by decreasing  $p_T$ . Note that the top and electroweak backgrounds are subdominant and are not shown here.

example, in figure 1 where the signal event has  $T_{43} = 0.45$  compared to 0.73 for the QCD event. As figure 5 shows, after a cut on the total jet mass (right panel), the ratio of signal to background improves relative to no total jet mass cut (left panel). The right panel suggests that the signal and background can be distinguished by applying an additional cut  $T_{43} \lesssim 0.6$ . We demonstrate the efficacy of this strategy in the following section when we estimate the sensitivity to the signal topologies in figure 2.

### 3 Analysis strategy

Having presented the individual components of our analysis, we now combine them and present the complete search strategy. To illustrate the effectiveness of this approach, we compute expected limits for the three different RPV gluino decay chains in figure 2. Of course, our proposal is quite general and can be applied to any high-multiplicity final-state.



**Figure 5.** Distributions of  $T_{43}$  for backgrounds and an example signal, with  $M_J > 0$  (left) and  $M_J > 500$  GeV (right). The signal (red solid line) is pair production of a 750 GeV gluino with  $\tilde{g} \rightarrow t\bar{t} + 3j$ . The stacked histogram is for background (QCD in solid blue,  $W^\pm/Z^0 + 4j$  in hatched magenta, and  $t\bar{t} + j$  in striped green). A cut on  $T_{43} \lesssim 0.6$  helps to distinguish signal from background, after requiring  $M_J > 500$  GeV.

### 3.1 Event generation

We begin by briefly describing the generation of signal and background events. Appendix A contains a more detailed description of the detector mockup and Monte Carlo validation.

QCD is the dominant background for a multijet signal with no missing energy. **Sherpa** 1.4.0 [56–60] is used to generate and shower  $\sim 400$  million inclusive  $pp \rightarrow nj$  events, where  $n \in (2, \dots, 6)$ . Matrix elements for up to 6 partons are generated, which are then matched to the parton shower using the CKKW procedure [61]. All Sherpa events are generated using the default CTEQ 6.6 parton distribution function [62] and include the effects of underlying event. We generated a sample of weighted events in order to increase the statistical power of our finite sample. The Monte Carlo error,  $\epsilon_{MC}$ , after cuts is

$$\epsilon_{MC} = \frac{\sqrt{\sum_i w_i^2}}{\sum_i w_i}, \tag{3.1}$$

where  $w_i$  is the weight of the  $i^{\text{th}}$  event in the sample. We verify that the Monte Carlo error is less than the systematic error for the signal regions of interest.

For consistency, **Sherpa** is also used to generate additional subleading background contributions. In particular, we generate  $\sim 25$  million matched and weighted  $t\bar{t} + n_t j$  events, where the tops are forced to decay hadronically. We also simulate  $\sim 25$  million matched and weighted data sets for each electroweak background:  $W^+ + n_W j$ ,  $W^- + n_W j$ , and  $Z^0 + n_Z j$ , where the gauge bosons are forced to decay to quarks. Here,  $n_t \in 0, 1$  and  $n_W, n_Z \in 1, 2, 3, 4$ . Table 1 shows that these non-QCD backgrounds are subdominant. This would not be the case if a  $\cancel{E}_T$  cut were also applied.

The matrix elements for gluino pair production are generated in **MadGraph5** 1.4.8.4 [63] for the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology. Those for the  $\tilde{g} \rightarrow t + 2j$  and  $\tilde{g} \rightarrow 3j$  topologies are

Requirement	QCD	$t\bar{t} + j$	$W^\pm/Z^0 + 4j$	$\tilde{g} \rightarrow t\bar{t} + 3j$	$\tilde{g} \rightarrow t + 2j$	$\tilde{g} \rightarrow 3j$
(1) $N_j = 4$	$5.8 \times 10^6$	4500	$1.0 \times 10^4$	680	7200	4800
(2) $M_J > 500 \text{ GeV}$	6800	8.4	40	400	990	640
(3) $T_{43} < 0.6$ <i>or</i>	180	0.61	1.5	<b>75</b>	<b>110</b>	(48)
$T_{21} < 0.2$	77	0.047	1.1	(1.7)	(27)	<b>39</b>

**Table 1.** Event yields from our Monte Carlo simulation, assuming  $5 \text{ fb}^{-1}$  of 8 TeV data and taking the gluino mass to be 750 GeV for  $\tilde{g} \rightarrow t\bar{t} + 3j$  and 500 GeV for the other two topologies. The table shows the number of events after requiring (1) four fat jets with  $m_j > 20 \text{ GeV}$  and the appropriate  $p_T$  requirements (see text), then (2) a cut on the total jet mass, and then (3) a cut on event-subjettiness for a given choice of  $T_{NM}$ . Yields are shown for two different  $T_{NM}$  cuts that are optimized for the 18, 10, and 6-parton topologies; the number of events that corresponds to the best choice for this cut is bolded while the non-optimal choices are in parentheses.

generated directly in `Pythia` 8.170 [64–66], where the RPV gluino is allowed to hadronize before decaying. All three signal topologies are generated using the default CTEQ6L1 PDF set [67, 68] and are showered and hadronized in `Pythia` including the effects of underlying event. Because the gluinos are produced at threshold and decay to several fairly hard jets, it is not necessary to perform matching.

Both signal and background events are passed through our own detector mockup, which only includes the effects of detector granularity. `FastJet` 3.0 [69, 70] is used to cluster events into anti- $k_T$  [39] jets with  $R = 1.2$ . Variables such as jet mass and substructure are sensitive to soft, diffuse radiation that results from underlying event and pile-up. The ATLAS study in [33] explicitly demonstrated that the mean jet mass for anti- $k_T$  jets with  $R = 1.0$  and  $p_T > 300 \text{ GeV}$  is constant with respect to the number of pile-up vertices for  $35 \text{ pb}^{-1}$  of 7 TeV data, after a splitting/filtering procedure is applied. For variable multiplicity fat jets, which is quite typical for accidental substructure, filtering is not the optimal grooming technique because it places a fixed requirement on the number of subjets within the fat jet [4]. Instead, to reduce the contamination due to soft radiation resulting from underlying event, we apply the trimming procedure of [11] to the jets before applying any kinematic cuts. We require any subjets of radius  $R = 0.3$  to have a  $p_T$  greater than 5% of the fat jet’s transverse momentum. This choice of parameters is motivated by a recent ATLAS analysis [32]. We find that trimming eliminates the dependence on the different underlying event models used by the generators.

`Prospino` 2.1 [71] is used to obtain the NLO production cross section for the gluinos. For the QCD background, we use a  $K$ -factor of 1.8, obtained by comparing distributions of the generated QCD Monte Carlo with published distributions in [33, 72] (see appendix A for details on validation). All other backgrounds are subdominant and our analysis is therefore insensitive to the exact choice of their cross sections. We use the `Sherpa` leading order predictions for these backgrounds.

### 3.2 Expected reach

Now, we are ready to compute the expected reach of our analysis. All events are required to satisfy the following criteria. Each event must have at least four fat jets, where the

$p_T$  of the hardest jet is at least 100 GeV and the  $p_T$  of the next three hardest jets is at least 50 GeV. To reduce contamination of heavy flavor resonances and high- $p_T$  QCD jets with no hard splittings, only jets with  $m_j > 20$  GeV are considered. To further reduce QCD and  $t\bar{t}$  background contributions, each event must have at least 500 GeV of total jet mass,  $M_J$ . Finally, a cut is placed on event-subjettiness,  $T_{NM}$ . The cuts for  $M_J$  and  $T_{NM}$  were selected to maximize significance, while ensuring that the Monte Carlo error remained below the systematic error. This requirement imposes a significant limitation on our ability to fully optimize the search and is the reason we only present one set of cuts. Table 1 summarizes the cut efficiencies on signal and background.

To determine the expected reach for the three topologies in figure 2, we assume that the probability of measuring  $n$  events is given by the Poisson distribution with mean  $\mu = B + S$ , where  $B$  and  $S$  are the number of expected background and signal events, respectively. The probability of measuring up to  $N_m$  events is

$$P(N_m|\mu) = e^{-\mu} \sum_{n=0}^{N_m} \frac{\mu^n}{n!}. \tag{3.2}$$

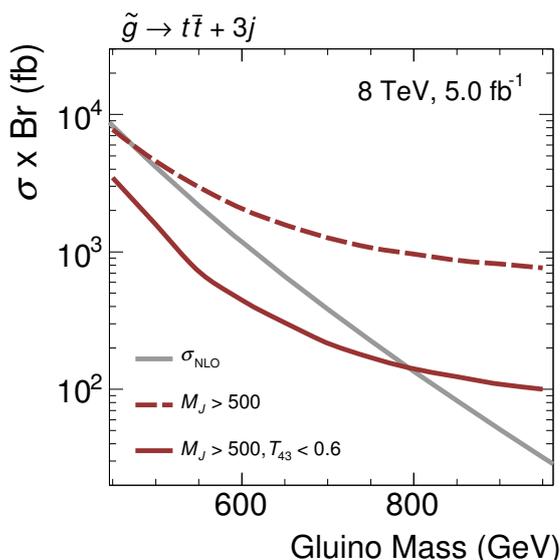
This expression assumes that there is no uncertainty in the value for  $B$ . In the presence of a systematic uncertainty  $\epsilon_{\text{sys}}$ , eq. (3.2) must be convoluted with the probability distribution of  $B$ , which we assume is log-normal because  $B \geq 0$ :

$$P_{\text{sys}}(N_m, S, B) = \int_0^\infty dx P(N_m|S+x) \cdot \ln \mathcal{N}(x), \tag{3.3}$$

where  $\ln \mathcal{N}(x) = \frac{1}{x\sqrt{2\pi\epsilon_{\text{sys}}}} \exp\left[-\frac{(\ln x - \ln B)^2}{2\epsilon_{\text{sys}}^2}\right]$ . Note that as  $\epsilon_{\text{sys}} \rightarrow 0$ , the log-normal distribution becomes a delta function centered at  $B$  and eq. (3.3) reproduces the standard result for a Poisson distribution. To obtain the expected limit on the signal cross section, we solve eq. (3.3) for  $S$  assuming that  $N_m = B$  and  $P_{\text{sys}} = 0.05$  (95% exclusion). We find that the expected limits are not sensitive to the distribution function chosen for  $B$ ; a Gaussian distribution gives essentially the same result.

An ATLAS analysis of the full 2011 dataset reported a jet mass scale systematic uncertainty of  $\sim 4\text{--}8\%$  (depending on jet  $p_T$ ) for anti- $k_T$  trimmed jets with  $R = 1.0$  [32]. For four fat jets, this gives at most a 16% systematic uncertainty. To be conservative and to account for additional sources of systematic effects (e.g. jet energy scale), we take  $\epsilon_{\text{sys}} = 20\%$  when computing sensitivities.

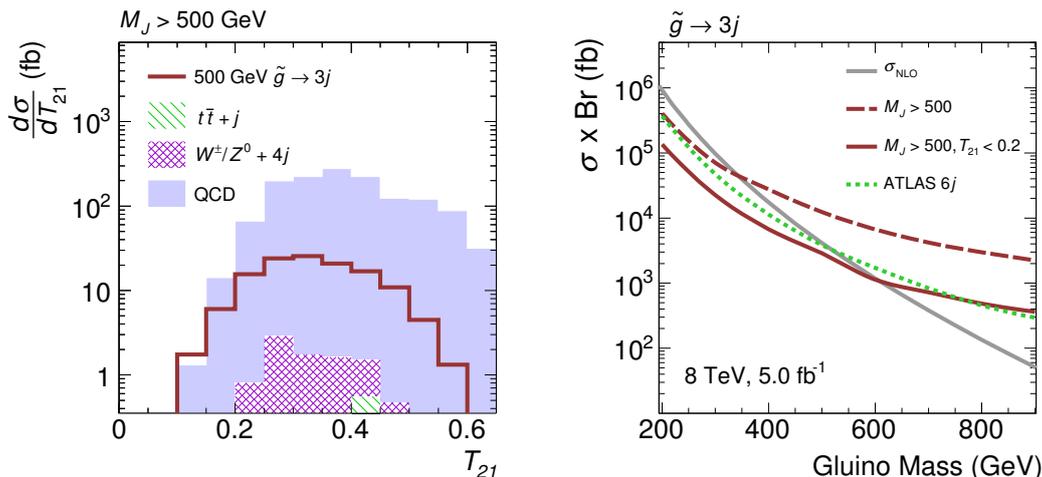
We begin by considering gluino pair production with  $\tilde{g} \rightarrow t\bar{t} + 3j$ . This topology can yield up to 18 partons when the tops decay hadronically. For this final state, the  $T_{43}$  event-subjettiness variable is most effective. For a 750 GeV gluino, a cut of  $T_{43} < 0.6$  increases  $S/B$  from 0.06 to 0.42, and  $S/\sqrt{B}$  from 4.9 to 5.6 as seen in table 1. Figure 6 shows the expected reach for  $5 \text{ fb}^{-1}$  of 8 TeV data. The gray line is the NLO gluino pair-production cross section, as evaluated by **Prospino**. The dashed red line shows the expected limit when all cuts are applied, except that on event-subjettiness. With the additional cut on  $T_{43}$ , the expected limit improves by  $\sim 350$  GeV, as illustrated by the solid red line. Requiring jets with accidental substructure significantly extends the reach beyond a search that relies on total jet mass alone.



**Figure 6.** The 95% expected exclusion curves for the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology at the 8 TeV LHC with  $5 \text{ fb}^{-1}$  of data. The solid grey curve is the NLO prediction for the gluino pair production cross section computed using *Prospino*, the dashed red curve is the expected exclusion including all cuts except the one on event-subjettiness, and the solid red curve is the exclusion when  $T_{43} < 0.6$  is imposed. A systematic error  $\epsilon_{\text{sys}} = 20\%$  is assumed for the background prediction. Cutting on event-subjettiness improves the reach by  $\sim 350$  GeV.

Event-subjettiness is an effective variable for other RPV gluino decay chains. However, as the number of hard partons decreases, the signature of accidental substructure becomes more subtle. Consider the middle diagram of figure 2 where  $\tilde{g} \rightarrow t + 2j$ . The 8 TeV,  $5 \text{ fb}^{-1}$  expected limits on this final state are extended from 400 GeV to 600 GeV when  $T_{43} < 0.6$  is required in addition to a jet mass cut. For a 500 GeV gluino, cutting on substructure improves the signal to background ratio from 0.14 to 0.61 as seen in table 1. Due to the smaller number of partons, the improvement in significance is not as dramatic as for the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology described previously. Here, the main advantage of cutting on substructure is to increase  $S/B$ . This provides a significant improvement because systematic uncertainties tend to drive the sensitivity in the signal region when QCD is the dominant background.

Lastly, we consider the 6-parton topology illustrated in the right-most diagram of figure 2. Of the three decay modes studied in this work, this has the fewest partons and is therefore the most challenging to observe. In particular,  $T_{21}$  provides the best discriminating power for this topology. The left panel of figure 7 shows the  $T_{21}$  distribution for background and signal after applying a  $M_J > 500$  GeV cut. The background is peaked between 0.35–0.4 and the signal is peaked at 0.25–0.35. The right panel of figure 7 shows the expected exclusion for the 6-parton final state, assuming  $5 \text{ fb}^{-1}$  of 8 TeV data. The dashed red line shows that the expected limit is  $\sim 350$  GeV before a cut on event-subjettiness. The expected limit increases to  $\sim 600$  GeV when  $T_{21} < 0.2$  is required (the solid red line). As in the last example, the improvement in the limit arises from an increase in the ratio of signal to background after substructure cuts.



**Figure 7.** The  $T_{21}$  distribution for signal and background after requiring  $M_J > 500$  GeV (left) and 95% expected exclusion (right) for the  $\tilde{g} \rightarrow 3j$  topology at the 8 TeV LHC with  $5 \text{ fb}^{-1}$  of data. *Left:* the signal (red solid line) is pair production of a 500 GeV gluino with  $\tilde{g} \rightarrow 3j$ . The stacked histogram is for background (QCD in solid blue,  $W^\pm/Z^0 + 4j$  in hatched magenta, and  $t\bar{t} + j$  in striped green). A cut on  $T_{21} \lesssim 0.2$  effectively distinguishes signal from background, after requiring  $M_J > 500$  GeV. *Right:* the solid grey curve is the NLO prediction for the gluino pair production cross section computed using *Prospino*, the dashed red curve is the expected exclusion including all cuts except the one on event-subjettiness, and the solid red curve is the exclusion when  $T_{21} < 0.2$  is imposed. For comparison, the green dotted line shows our reproduction of the ATLAS search for this same topology [37]. Our analysis is competitive with the ATLAS reach. A systematic error  $\epsilon_{\text{sys}} = 20\%$  is assumed for the background prediction. A cut on event-subjettiness improves the reach by  $\sim 250$  GeV.

The expected reach of our substructure analysis for RPV gluinos is  $\sim 600$  GeV and compares favorably with published limits from CMS and ATLAS. The CMS search for three-jet invariant mass resonances [40] excludes an RPV gluino from 280–460 GeV with  $5 \text{ fb}^{-1}$  of 7 TeV data. The ATLAS analysis for this final state, published with  $4.6 \text{ fb}^{-1}$  of 7 TeV data, uses two techniques to provide exclusions [37]. They perform a boosted gluino analysis that makes use of jet substructure and can exclude the gluino in the range 100–255 GeV.<sup>4</sup> A separate “resolved” analysis uses the  $p_T$  of the sixth jet (anti- $k_T$ ,  $R = 0.4$ ) to separate signal from background, and excludes the gluino from 100–666 GeV.

To provide a direct comparison, we reproduce the ATLAS resolved analysis by reclustering our background and signal into anti- $k_T$  jets with  $R = 0.4$  and applying the cuts from [37]. The projected limit for  $5 \text{ fb}^{-1}$  of 8 TeV data is shown by the green dotted line of figure 7 and gives a limit of about 550 GeV.<sup>5</sup> This demonstrates that our projected limit, which relies on accidental substructure is competitive to that from the ATLAS resolved analysis.

<sup>4</sup>The recent theory work in [38] finds that the limit on boosted RPV gluinos can be increased by searching for a peak in the jet mass spectrum.

<sup>5</sup>Note that our expected limit of 550 GeV is weaker than that in [37], although it does fall at the edge of the published 1-sigma uncertainty. We can reproduce their limit if we take a  $K$ -factor of 1.0 for the QCD background. For consistency with the validation plots from appendix A, we use the more conservative 1.8  $K$ -factor for figure 7.

To emphasize the effectiveness of our approach, we also performed a naive comparison between our method and the ATLAS resolved jet analysis of [37] as applied to the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology. The ATLAS search is not optimized for this signal; in particular, for this topology, relying on  $b$ -jets and/or leptons may be a more effective strategy. However, it provides a rough guide for a small-radius jet (with  $R \sim 0.4$ ) analysis that one might consider when searching for this multijet topology. We find that there is no bound on the gluino mass for the 6-jet cuts proposed in [37]. In principle, the signal region could be extended to a larger jet count. In that case, however, background estimation can be quite challenging. On the other hand, the accidental substructure analysis outlined in this paper is broadly applicable to signals with different jet multiplicities.

## 4 Conclusion

In this paper, we introduced the concept of accidental substructure and illustrated its usefulness in searches for high-multiplicity final states and no missing energy. Accidental substructure arises because there is a high likelihood that several final-state partons will be clustered together in the same large-radius jet. These final state partons need not have originated from the same parent particle. QCD is the dominant background. Having several partons in a QCD event that undergo a large-angle, hard splitting is rare enough to make accidental substructure a useful discriminator.

We analyzed three RPV gluino decay topologies with as many as 18, 10, and 6 partons in the final state. The requirement that the total jet mass be greater than 500 GeV, in conjunction with a cut on event-subjettiness, proved to be very effective. We found projected limits of  $\mathcal{O}(800 \text{ GeV})$  for the  $\tilde{g} \rightarrow t\bar{t} + 3j$  topology,  $\mathcal{O}(600 \text{ GeV})$  for the  $\tilde{g} \rightarrow t + 2j$  topology, and  $\mathcal{O}(600 \text{ GeV})$  for the  $\tilde{g} \rightarrow 3j$  final state with  $5 \text{ fb}^{-1}$  of 8 TeV data. These projections assume a 20% systematic uncertainty and a conservative  $K$ -factor for the normalization of the QCD background. Our goal was to illustrate the general applicability of a search using accidental substructure and we expect that many aspects of this analysis can be further optimized. One possibility, for instance, is to use a neural network to select the appropriate  $N$ -subjettiness variables to include in the evaluation of event-subjettiness. Also, we have not explored how the sensitivity of the search depends on jet radius.

In the case of the 6-parton final state from RPV gluino decays, our expected limit is comparable to that set by the ATLAS small-radius jet analysis [37]. Determining the normalization of the QCD background for a 6 (or more) small-radius jet signal is challenging. As a result, it is important to have a complementary search with independent systematics. Our accidental substructure search is one possible example and is, in addition, sensitive to a broader array of signals than the ATLAS search. In particular, its sensitivity only improves as the number of final-state partons increases, as we showed for the 10 and 18-parton final states.

Events with many tops can lead to many jets in the final state (the scenario we consider here), but other decay channels can give leptons and  $\cancel{E}_T$ . Analyses that tag on a lepton and several  $b$ -jets can be sensitive in these cases [73]. We also expect our reach to improve significantly when  $b$ -tags are included [74]. Alternatively, the total energy  $S_T$  may be useful;

while it provides the greatest discriminating power in black hole searches [75, 76], the  $S_T$  cut must be above several TeV to adequately reduce the multijet background. Tagging on a lepton in addition to six or more jets, could allow an  $S_T$  cut down to  $\sim 1$  TeV [77].

The search we proposed here is complementary to these types of analyses. We expect that its potential reach will only increase by adding additional handles. For example, we find that naive cuts on jet mass and event-subjettiness lead to a limit on  $\tilde{g} \rightarrow t\bar{t} + \cancel{E}_T$  that is only slightly weaker than the current bounds from CMS and ATLAS. Adding a lepton, a  $b$ -tag and/or a small cut on  $\cancel{E}_T$  could make the search even more powerful.

A significant advantage of using fat jets to study final states with many partons is that it is compatible with data-driven determinations of the QCD background. Mapping out the phase-space of high multiplicity QCD with Monte Carlo is currently not possible. For a fat jet analysis, one can use a dijet sample to map out distributions of the internal structure of the jets and to obtain templates for jet mass and substructure as a function of the jet kinematics. Under the mild assumption that the correlations between fat jets are small, one only needs to predict the phase space distribution of the four fat jets, while the internal properties of each fat jet can be modeled using the template functions derived from the dijet events. This simple algorithm allows an extrapolation of the QCD contribution to the four fat jet signal region.

The possibility of using a jet's internal structure to learn about its origin provides exciting opportunities for new physics searches at the LHC. Although jet substructure has only been used for boosted signals thus far, this work demonstrates that it is also applicable in the non-boosted regime. We have shown that accidental substructure provides a robust and powerful new paradigm for new physics searches at the LHC, complementing and extending the reach of current analyses.

**Note added.** A related work will appear [78], which proposes a method of subjet counting and applies it to searches for high-multiplicity signals.

## Acknowledgments

We thank R. D'Agnolo, D. Curtin, S. El-Hedri, R. Essig, M. Jankowiak, D. Miller, M. Pierini, B. Shuve, and B. Tweedie for useful discussions. We are especially thankful to S. Hoeche for his help with the QCD Sherpa generation, P. Skands for his help with `Pythia 8`, and J. Wacker for many conversations on data-driven approaches. TC is supported by the US Department of Energy under contract number DE-AC02-76SF00515. EI is supported in part by the Government of Canada through Industry Canada, by the Province of Ontario through the Ministry of Research and Information, by the US Department of Energy under contract number DE-AC02-76SF00515, and by an LHCTI graduate fellowship under grant NSF-PHY-0969510. ML is supported by the Simons Postdoctoral Fellows Program. HKL is supported by the DOE Office of Science Graduate Fellowship Program. TC and ML would like to thank the Aspen Center for Physics, which is supported by the National Science Foundation grant number 1066293. ML also acknowledges the Galileo Galilei Institute for Theoretical Physics for their hospitality.

## A Simulation details and validation

In this appendix, we discuss the details of our simple detector mockup and provide validation plots comparing our QCD Monte Carlo to a number of public distributions from ATLAS. We extract a  $K$ -factor to normalize our QCD sample and demonstrate that our Monte Carlo reproduces the measured shapes of substructure and jet mass distributions to reasonable accuracy.

We simulate detector granularity by clustering stable, visible generator-level particles into  $\eta \times \phi$  cells of size  $0.1 \times 0.1$ . Electrons, muons, and photons are kept if they fall within  $|\eta| < 2.5$ , while all other particles are kept if they fall within  $|\eta| < 3.0$ . Each calorimeter cell is assigned a light-like vector with energy equal to the sum of all particle energies contained therein. `FastJet 3.0` clusters these four-vectors into anti- $k_T$  jets and computes  $N$ -subjettiness for the resulting jets using the “min\_axes” algorithm, implemented in the  $N$ -subjettiness plugin of Thaler and Van-Tilberg [14, 19]. Note that leptons are included in jet clustering and when calculating substructure variables. A jet is removed if it is within  $\Delta R < 0.2$  of a lepton and its  $p_T$  is less than twice the lepton’s  $p_T$ .

We validate our QCD Monte Carlo by comparing against published kinematic and substructure distributions. No published 8 TeV substructure results are currently available, and so we compare against the published 7 TeV ATLAS results [33, 37, 72]. A weighted sample of  $pp \rightarrow nj$ , where  $n \in (2, \dots, 6)$ , is generated in `Sherpa 1.4.0`. Our 7 TeV sample consists of  $\sim 50$  million events and is generated with the same settings as our  $\sim 400$  million event 8 TeV Sherpa sample, described in section 3.1.

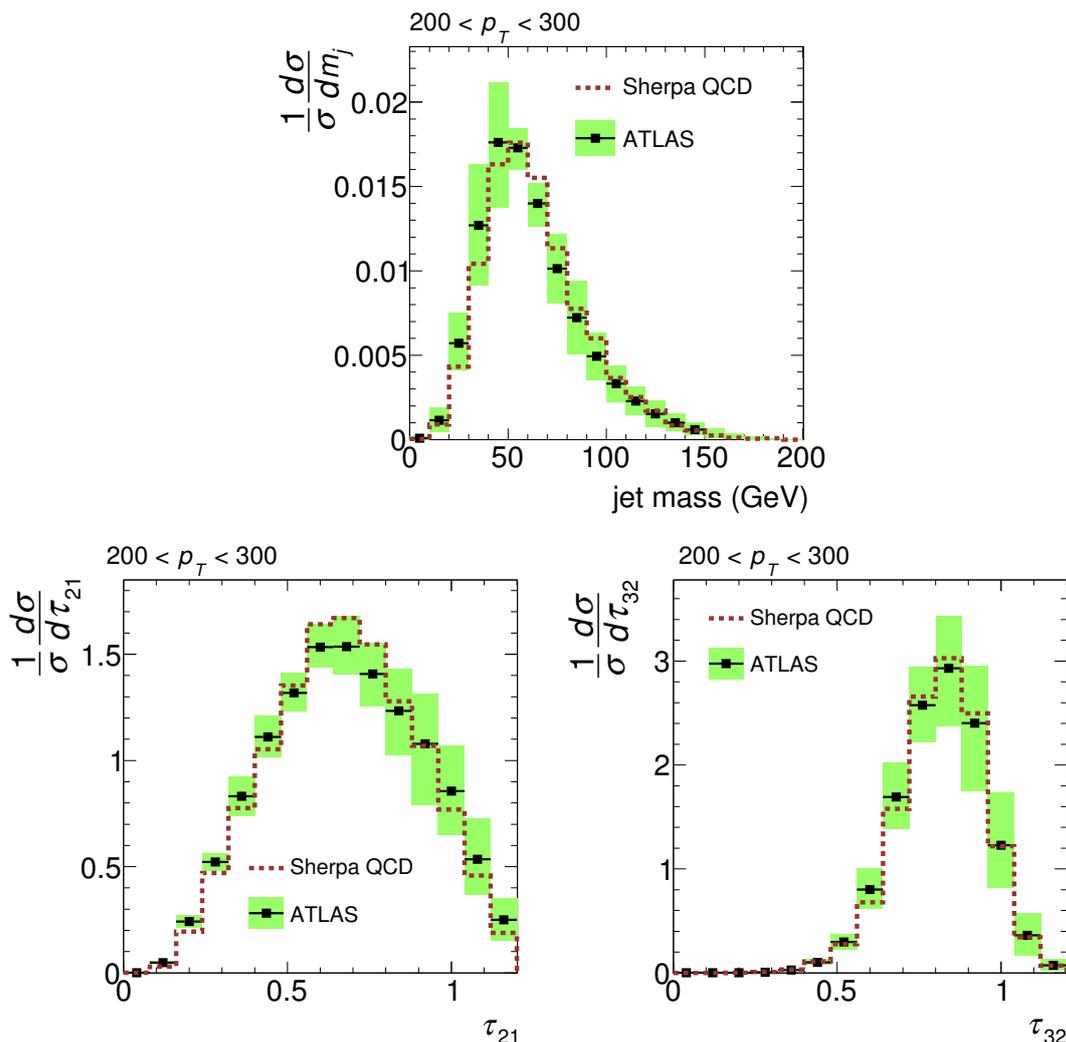
To validate the shape of the jet mass and substructure distributions, we follow the analysis in [33] and compare to the unfolded distributions. Particles are clustered into anti- $k_T$  jets with  $R = 1.0$ . The resulting jets are divided into four equally-spaced  $p_T$  bins from 200 to 600 GeV. The jet mass ( $\tau_{21}$  and  $\tau_{32}$ ) distributions are shown in the top (bottom) of figure 8 for  $p_T \in (200, 300)$ . The Monte Carlo predictions are well within the error bands quoted by ATLAS. We checked that the `Sherpa` results for the higher  $p_T$  bins, not shown here, also match the ATLAS results.

`Sherpa` outputs a leading order (matched) cross section of  $\sigma_{\text{QCD}}^{\text{Sherpa}} = 9.6 \times 10^9$  fb. Because this cross section is enhanced by loop effects, we must find the proper normalization, or  $K$ -factor, for the QCD background:

$$\sigma_{\text{QCD}} = K \times \sigma_{\text{QCD}}^{\text{Sherpa}}. \quad (\text{A.1})$$

Using the reported 2-jet inclusive cross-section in [72], we obtain a  $K$ -factor of  $\sim 1.3$ . Comparing to the 6<sup>th</sup> jet  $p_T$  distribution in [37], we obtain a  $K$ -factor of 1.8. Furthermore, by comparing the normalization of the jet mass,  $\tau_{21}$  and  $\tau_{32}$  distributions in [33] we obtain a  $K$ -factor of 1.8. To be conservative, we assume a  $K$ -factor of 1.8 in this work.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.



**Figure 8.** Jet mass (top) and  $N$ -subjettiness (bottom) comparisons between the Sherpa QCD prediction [dotted red] and the ATLAS results [black rectangle] of [33]. The green band is the combined statistical and systematic error in the ATLAS measurement including the uncertainty from the unfolding procedure.

## References

- [1] G.P. Salam, *Towards jetography*, *Eur. Phys. J. C* **67** (2010) 637 [[arXiv:0906.1833](#)] [[INSPIRE](#)].
- [2] A. Abdesselam et al., *Boosted objects: a probe of beyond the standard model physics*, *Eur. Phys. J. C* **71** (2011) 1661 [[arXiv:1012.5412](#)] [[INSPIRE](#)].
- [3] A. Altheimer et al., *Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*, *J. Phys. G* **39** (2012) 063001 [[arXiv:1201.0008](#)] [[INSPIRE](#)].
- [4] 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](#)].

- [5] J. Butterworth, J.R. Ellis and A. Raklev, *Reconstructing sparticle mass spectra using hadronic decays*, *JHEP* **05** (2007) 033 [[hep-ph/0702150](#)] [[INSPIRE](#)].
- [6] J. Thaler and L.-T. Wang, *Strategies to identify boosted tops*, *JHEP* **07** (2008) 092 [[arXiv:0806.0023](#)] [[INSPIRE](#)].
- [7] L.G. Almeida, S.J. Lee, G. Perez, I. Sung and J. Virzi, *Top jets at the LHC*, *Phys. Rev. D* **79** (2009) 074012 [[arXiv:0810.0934](#)] [[INSPIRE](#)].
- [8] D.E. Kaplan, K. Rehermann, M.D. Schwartz and B. Tweedie, *Top tagging: a method for identifying boosted hadronically decaying top quarks*, *Phys. Rev. Lett.* **101** (2008) 142001 [[arXiv:0806.0848](#)] [[INSPIRE](#)].
- [9] 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](#)].
- [10] T. Plehn, G.P. Salam and M. Spannowsky, *Fat jets for a light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [[arXiv:0910.5472](#)] [[INSPIRE](#)].
- [11] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, *JHEP* **02** (2010) 084 [[arXiv:0912.1342](#)] [[INSPIRE](#)].
- [12] G.D. Kribs, A. Martin, T.S. Roy and M. Spannowsky, *Discovering the Higgs boson in new physics events using jet substructure*, *Phys. Rev. D* **81** (2010) 111501 [[arXiv:0912.4731](#)] [[INSPIRE](#)].
- [13] T. Plehn and M. Spannowsky, *Top tagging*, *J. Phys. G* **39** (2012) 083001 [[arXiv:1112.4441](#)] [[INSPIRE](#)].
- [14] J. Thaler and K. Van Tilburg, *Identifying boosted objects with  $N$ -subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [15] J. Gallicchio and M.D. Schwartz, *Seeing in color: jet superstructure*, *Phys. Rev. Lett.* **105** (2010) 022001 [[arXiv:1001.5027](#)] [[INSPIRE](#)].
- [16] J. Gallicchio et al., *Multivariate discrimination and the Higgs +  $W/Z$  search*, *JHEP* **04** (2011) 069 [[arXiv:1010.3698](#)] [[INSPIRE](#)].
- [17] Y. Cui, Z. Han and M.D. Schwartz,  *$W$ -jet tagging: optimizing the identification of boosted hadronically-decaying  $W$  bosons*, *Phys. Rev. D* **83** (2011) 074023 [[arXiv:1012.2077](#)] [[INSPIRE](#)].
- [18] L.G. Almeida, S.J. Lee, G. Perez, G. Sterman and I. Sung, *Template overlap method for massive jets*, *Phys. Rev. D* **82** (2010) 054034 [[arXiv:1006.2035](#)] [[INSPIRE](#)].
- [19] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing  $N$ -subjettiness*, *JHEP* **02** (2012) 093 [[arXiv:1108.2701](#)] [[INSPIRE](#)].
- [20] A. Hook, M. Jankowiak and J.G. Wacker, *Jet dipolarity: top tagging with color flow*, *JHEP* **04** (2012) 007 [[arXiv:1102.1012](#)] [[INSPIRE](#)].
- [21] J. Gallicchio and M.D. Schwartz, *Quark and gluon tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [[arXiv:1106.3076](#)] [[INSPIRE](#)].
- [22] D.E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys. Rev. D* **84** (2011) 074002 [[arXiv:1102.3480](#)] [[INSPIRE](#)].
- [23] M. Jankowiak and A.J. Larkoski, *Jet substructure without trees*, *JHEP* **06** (2011) 057 [[arXiv:1104.1646](#)] [[INSPIRE](#)].

- [24] S.D. Ellis, A. Hornig, T.S. Roy, D. Krohn and M.D. Schwartz, *Qjets: a non-deterministic approach to tree-based jet substructure*, *Phys. Rev. Lett.* **108** (2012) 182003 [[arXiv:1201.1914](#)] [[INSPIRE](#)].
- [25] M. Jankowiak and A.J. Larkoski, *Angular scaling in jets*, *JHEP* **04** (2012) 039 [[arXiv:1201.2688](#)] [[INSPIRE](#)].
- [26] Z. Han, A. Katz, M. Son and B. Tweedie, *Boosting searches for natural SUSY with RPV via gluino cascades*, [arXiv:1211.4025](#) [[INSPIRE](#)].
- [27] D.E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, [arXiv:1211.3140](#) [[INSPIRE](#)].
- [28] D0 collaboration, V.M. Abazov et al., *Measurement of color flow in  $t\bar{t}$  events from  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Rev. D* **83** (2011) 092002 [[arXiv:1101.0648](#)] [[INSPIRE](#)].
- [29] CDF collaboration, T. Aaltonen et al., *Study of substructure of high transverse momentum jets produced in proton-antiproton collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Rev. D* **85** (2012) 091101 [[arXiv:1106.5952](#)] [[INSPIRE](#)].
- [30] ATLAS collaboration, *High  $p_T$  hadronic top quark identification*, *ATL-PHYS-CONF-2008-008* (2008).
- [31] ATLAS collaboration, *Measurement of jet mass and substructure for inclusive jets in  $\sqrt{s} = 7$  TeV  $pp$  collisions with the ATLAS experiment*, *ATLAS-CONF-2011-073* (2011).
- [32] ATLAS collaboration, *Performance of large- $R$  jets and jet substructure reconstruction with the ATLAS detector*, *ATLAS-CONF-2012-065* (2012).
- [33] ATLAS collaboration, *Jet mass and substructure of inclusive jets in  $\sqrt{s} = 7$  TeV  $pp$  collisions with the ATLAS experiment*, *JHEP* **05** (2012) 128 [[arXiv:1203.4606](#)] [[INSPIRE](#)].
- [34] ATLAS collaboration, *Study of jet shapes in inclusive jet production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *Phys. Rev. D* **83** (2011) 052003 [[arXiv:1101.0070](#)] [[INSPIRE](#)].
- [35] ATLAS collaboration, *ATLAS measurements of the properties of jets for boosted particle searches*, *Phys. Rev. D* **86** (2012) 072006 [[arXiv:1206.5369](#)] [[INSPIRE](#)].
- [36] CMS collaboration, *Shape, transverse size and charged hadron multiplicity of jets in  $pp$  collisions at 7 TeV*, *JHEP* **06** (2012) 160 [[arXiv:1204.3170](#)] [[INSPIRE](#)].
- [37] ATLAS collaboration, *Search for pair production of massive particles decaying into three quarks with the ATLAS detector in  $\sqrt{s} = 7$  TeV  $pp$  collisions at the LHC*, *JHEP* **12** (2012) 086 [[arXiv:1210.4813](#)] [[INSPIRE](#)].
- [38] D. Curtin, R. Essig and B. Shuve, *Boosted multijet resonances and new color-flow variables*, [arXiv:1210.5523](#) [[INSPIRE](#)].
- [39] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [40] CMS collaboration, *Search for three-jet resonances in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **718** (2012) 329 [[arXiv:1208.2931](#)] [[INSPIRE](#)].
- [41] CMS collaboration, *Search for multijet resonances in the 8-jet final state*, *CMS-PAS-EXO-11-075* (2011).
- [42] E. Gerwick, T. Plehn, S. Schumann and P. Schichtel, *Scaling patterns for QCD jets*, *JHEP* **10** (2012) 162 [[arXiv:1208.3676](#)] [[INSPIRE](#)].

- [43] E. Gerwick, B. Gripaios, S. Schumann and B. Webber, *QCD jet rates with the inclusive generalized  $k_t$  algorithms*, [arXiv:1212.5235](#) [INSPIRE].
- [44] J. Fan, M. Reece and J.T. Ruderman, *Stealth supersymmetry*, *JHEP* **11** (2011) 012 [[arXiv:1105.5135](#)] [INSPIRE].
- [45] J. Fan, M. Reece and J.T. Ruderman, *A stealth supersymmetry sampler*, *JHEP* **07** (2012) 196 [[arXiv:1201.4875](#)] [INSPIRE].
- [46] C. Kilic, T. Okui and R. Sundrum, *Vectorlike confinement at the LHC*, *JHEP* **02** (2010) 018 [[arXiv:0906.0577](#)] [INSPIRE].
- [47] C. Gross, G. Marques Tavares, M. Schmaltz and C. Spethmann, *Light axigluon explanation of the Tevatron  $t\bar{t}b\bar{a}$  asymmetry and multijet signals at the LHC*, *Phys. Rev. D* **87** (2013) 014004 [[arXiv:1209.6375](#)] [INSPIRE].
- [48] J. Kang and M.A. Luty, *Macroscopic strings and 'quirks' at colliders*, *JHEP* **11** (2009) 065 [[arXiv:0805.4642](#)] [INSPIRE].
- [49] K. Cheung, W.-Y. Keung and T.-C. Yuan, *Phenomenology of iquarkonium*, *Nucl. Phys. B* **811** (2009) 274 [[arXiv:0810.1524](#)] [INSPIRE].
- [50] R. Barbier et al., *R-parity violating supersymmetry*, *Phys. Rept.* **420** (2005) 1 [[hep-ph/0406039](#)] [INSPIRE].
- [51] C. Brust, A. Katz and R. Sundrum, *SUSY stops at a bump*, *JHEP* **08** (2012) 059 [[arXiv:1206.2353](#)] [INSPIRE].
- [52] CMS collaboration, *A search for the decays of a new heavy particle in multijet events with the razor variables at CMS in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, CMS-PAS-SUS-12-009 (2012).
- [53] A. Hook, E. Izaguirre, M. Lisanti and J.G. Wacker, *High multiplicity searches at the LHC using jet masses*, *Phys. Rev. D* **85** (2012) 055029 [[arXiv:1202.0558](#)] [INSPIRE].
- [54] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn,  *$N$ -jettiness: an inclusive event shape to veto jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [[arXiv:1004.2489](#)] [INSPIRE].
- [55] G. Soyez, G.P. Salam, J. Kim, S. Dutta and M. Cacciari, *Pileup subtraction for jet shapes*, [arXiv:1211.2811](#) [INSPIRE].
- [56] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007 [[arXiv:0811.4622](#)] [INSPIRE].
- [57] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: a matrix element generator in C++*, *JHEP* **02** (2002) 044 [[hep-ph/0109036](#)] [INSPIRE].
- [58] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038 [[arXiv:0709.1027](#)] [INSPIRE].
- [59] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039 [[arXiv:0808.3674](#)] [INSPIRE].
- [60] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053 [[arXiv:0903.1219](#)] [INSPIRE].
- [61] S. Catani, F. Krauss, R. Kuhn and B. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [[hep-ph/0109231](#)] [INSPIRE].
- [62] P.M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev. D* **78** (2008) 013004 [[arXiv:0802.0007](#)] [INSPIRE].

- [63] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: going beyond*, *JHEP* **06** (2011) 128 [[arXiv:1106.0522](#)] [[INSPIRE](#)].
- [64] T. Sjöstrand et al., *High-energy physics event generation with PYTHIA 6.1*, *Comput. Phys. Commun.* **135** (2001) 238 [[hep-ph/0010017](#)] [[INSPIRE](#)].
- [65] T. Sjöstrand and P.Z. Skands, *Baryon number violation and string topologies*, *Nucl. Phys. B* **659** (2003) 243 [[hep-ph/0212264](#)] [[INSPIRE](#)].
- [66] N. Desai and P.Z. Skands, *Supersymmetry and generic BSM models in PYTHIA 8*, *Eur. Phys. J. C* **72** (2012) 2238 [[arXiv:1109.5852](#)] [[INSPIRE](#)].
- [67] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012 [[hep-ph/0201195](#)] [[INSPIRE](#)].
- [68] D. Stump et al., *Inclusive jet production, parton distributions and the search for new physics*, *JHEP* **10** (2003) 046 [[hep-ph/0303013](#)] [[INSPIRE](#)].
- [69] 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](#)].
- [70] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [71] W. Beenakker, R. Hopker, M. Spira and P. Zerwas, *Squark and gluino production at hadron colliders*, *Nucl. Phys. B* **492** (1997) 51 [[hep-ph/9610490](#)] [[INSPIRE](#)].
- [72] ATLAS collaboration, *Measurement of multi-jet cross sections in proton-proton collisions at a 7 TeV center-of-mass energy*, *Eur. Phys. J. C* **71** (2011) 1763 [[arXiv:1107.2092](#)] [[INSPIRE](#)].
- [73] CMS collaboration, *Search for heavy quarks decaying into a top quark and a W or Z boson using lepton + jets events in pp collisions at  $\sqrt{s} = 7$  TeV*, *JHEP* **01** (2013) 154 [[arXiv:1210.7471](#)] [[INSPIRE](#)].
- [74] CMS collaboration, *Identification of b-quark jets with the CMS experiment*, [arXiv:1211.4462](#) [[INSPIRE](#)].
- [75] CMS collaboration, *Search for microscopic black holes in pp collisions at  $\sqrt{s} = 7$  TeV*, *JHEP* **04** (2012) 061 [[arXiv:1202.6396](#)] [[INSPIRE](#)].
- [76] ATLAS collaboration, *Search for TeV-scale gravity signatures in final states with leptons and jets with the ATLAS detector at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **716** (2012) 122 [[arXiv:1204.4646](#)] [[INSPIRE](#)].
- [77] M. Lisanti, P. Schuster, M. Strassler and N. Toro, *Study of LHC searches for a lepton and many jets*, *JHEP* **11** (2012) 081 [[arXiv:1107.5055](#)] [[INSPIRE](#)].
- [78] S.E. Hedri, A. Hook, M. Jankowiak and J.G. Wacker, *Learning how to count: a high multiplicity search for the LHC*, [arXiv:1302.1870](#) [[INSPIRE](#)].