Available on CMS information server

**CMS IN 2008/XXX**

# CMS Internal Note

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

**3 November 2008**

# Exploiting the Kinematics of n-jet Topologies for Early SUSY Searches in the All-Hadronic Channel.

O. Buchmuller

*CERN, Geneva, Switzerland*

M. Stoye, T. Whyntie

*High Energy Physics Group, Imperial College London, United Kingdom*

**Abstract**

Building on the potential of the hadronic dijet SUSY search analysis, we present a possible strategy for extending the search to topologies with $n$ hadronic jets. In doing so we generalise the application of the $\alpha_T$ observable by clustering the $n$ jets to a "pseudo-dijet" system, using a jet merging scheme based on keeping the observed transverse mass the same for any combination of jets and the principle of balancing the pseudo-dijet transverse energies. The method predicts 2329.9 LM1 SUSY events for $n = 2 - 6$ for $1\text{fb}^{-1}$ of data, a five-fold increase on the dijet analysis, with $S/B = 7.0$. Suggestions are made for other $\alpha_T$-like variables based on similar principles and initial, unoptimised results are presented. The "self-protecting" nature of some of the observables used is briefly discussed, and future studies to investigate these properties are outlined.

# 1 Introduction

Any claimed discovery of non-Standard Model (SM) physics at the Large Hadron Collider must be robust to the uncertainties inherent in early machine data. Issues such as

- the poorly understood nature of SM background processes in the new energy regime;
- detector mismeasurements due to miscalibration and noise

should be anticipated and catered for by initial search strategies. This is particularly true of theories like supersymmetry where massive, weakly-interacting particles are produced in $pp$ collisions and escape detection.

Thus, in order to be able to annouce the discovery of new physics with such a signature with early data, it is useful to construct observables that offer some degree of "protection" against mismeasurements while maximising the signal yield for the low energies/luminosities of early machine running. In the following note we discuss a class of observables that appear to demonstrate such properties, following the work in the hadronic dijet channel reported in [2], [1]. We examine the kinematic properties and features of these variables for systems featuring multiple hadronic jets, where the dominant and least understood SM background at start-up will be multi-jet QCD events. We then present the results of some preliminary applications of these discriminants to hadronic systems with $n$-jets using the Monte Carlo datasets featured in [1], before outlining plans for future work.

It should be noted that, by definition, undetectable particles represent a mismeasurement of the system. We therefore distinguish between the "real missing energy" introduced to the system due to physical particles (e.g. neutrinos, LSPs) and "fake missing energy" caused by deficiencies in the detector. Backgrounds processes that have real missing energy, and cannot be eliminated by any other method, must be removed or constrained using data-driven methods. This is beyond the scope of this note.

# 2 Kinematics in the Transverse Plane

The $z$ momenta of the interacting partons in hadron-hadron collisions will vary in a way described by the relevant parton distribution function (PDF). While this is useful for probing a wide range of energies with a given machine configuration, it means that it is generally more useful to only consider the kinematics of an event in the transverse plane.

As such, we define the following variables for our system containing $n$ (hadronic) objects:

$$H_T = \sum_{i=1}^{n} E_{T(i)} \tag{1}$$

where, generally speaking, for final state particles $E_T = p_T$. This sets the scale of the interaction in the transverse plane.

$$\not{\mathbf{h}}_T = -\sum_{i=1}^{n} \mathbf{p}_{T(i)} = -\mathbf{h}_T \tag{2}$$

where $\mathbf{p}_T$ is the vectorial projection of the particle momentum in the transverse plane. This describes the scale and direction of the real and fake missing transverse energy of the system. We use a lower-case h to highlight the vectorial nature of $\not{\mathbf{h}}_T$ and $\mathbf{h}_T$.

$$M_{T(\mathrm{Tot.})} = \sqrt{H_T^2 - |\not{\mathbf{h}}_T|^2} \tag{3}$$

This represents the transverse mass of the system, having the form $E_T^2 - p_T^2$ when considered in terms of Equations 1 and 2. It is interesting to note that for a conserved, perfectly measured system, $M_T$ is maximal and equal to $H_T$, and that unlike $M$, the invariant mass of the system, $M_T$ is not invariant and will depend on the phase space available to the interaction and the kinematic configuration of the final state system.

# 3 The Dijet System - A Special Case

We now consider the hadronic dijet system featured in [2] and [1]. The variable used to reject most of the simulated QCD background after event preselection cuts, $\alpha_T$, is defined in [1] as

$$\alpha_T = \frac{E_{T(2)}}{M_{T(1,2)}}. \tag{4}$$

where the jets 1 and 2 are ordered by $E_T$. We can express this in terms of the variables defined in Section 2 by replacing the numerator with the following, equivalent, expression:

$$\alpha_T = \frac{\frac{1}{2}\left(E_{T(1)} + E_{T(2)} - \left|E_{T(1)} - E_{T(2)}\right|\right)}{M_{T(1,2)}} \tag{5}$$

$$= \frac{\frac{1}{2}\left(H_T - \Delta H_T\right)}{\sqrt{H_T^2 - |\not{H}_T|^2}} \tag{6}$$

where $\Delta H_T = |E_{T(1)} - E_{T(2)}|$ is the difference between the jet transverse energies. Thus for a perfectly balanced, perfectly measured dijet system, i.e. a QCD dijet event, $\Delta H_T = 0$ and $\not{H}_T = 0$ and so $\alpha_T = 0.5$. Likewise, events with large missing transverse energy with produce comparably small $M_{T(\text{Tot.})}$ values, and will be well above the cut regardless of the balance of the jet transverse energies.

In [1] the cut is made on $\alpha_T > 0.55$. The plot of $\alpha_T$ for the low mass SUSY point LM1 [3] and the dominant backgrounds (QCD, $Z \to \nu\nu$, $t\bar{t}$, $W+$ jets, etc.) for 1fb$^{-1}$ after the $H_T > 500$GeV preselection cut is shown in Fig. 1. Note the strong peak at $\alpha_T = 0.5$ for the QCD sample, and the large $\alpha_T$ values for the signal (and to a lesser extent, backgrounds with real $\not{H}_T$).

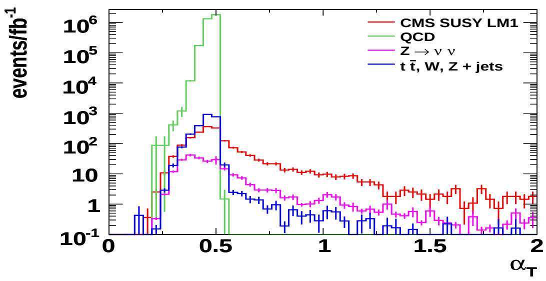


Figure 1: Plot of $\alpha_T$ for the hadronic dijet system after the preselection cuts up to and including $H_T >$ 500GeV, as outlined in [1].

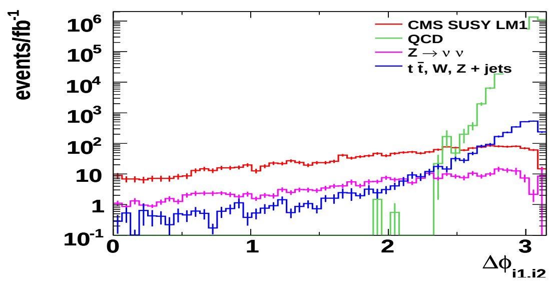


Figure 2: Plot of $\Delta\phi_{(1,2)}$ for the hadronic dijet system after the preselection cuts up to and including $H_T >$ 500GeV, as outlined in [1].

### 3.1 Other Kinematic Cuts in the Dijet Analysis

Another kinematic variable explored in [1] is $\Delta\phi\left(\mathbf{p}_{T(1)}, \mathbf{p}_{T(1)}\right) = \Delta\phi_{(1,2)}$, the angle between the two jets. Only events with

$$0 \leq \Delta\phi_{(1,2)} < \frac{2\pi}{3} \tag{7}$$

pass the cut, and so again back-to-back jets are rejected. Fig. 2 shows the plot of $\Delta\phi_{(1,2)}$ (again, for events passing the $H_T$ preselection cut, but not $\alpha_T$). In [1] this cut is applied after the $\alpha_T$ cut, and as can be seen from Table 1 it makes very little difference to the final signal and background yields. This suggests that the two variables contain very similar information for the dijet case; indeed, this relationship is explored in §5.1 of [1].

Another cut used in the dijet analysis compares the direction of $\not{H}_T$ with the direction of the third jet (if one with $E_T > 30$GeV is present). An event is rejected if

$$\Delta\phi(\not{H}_T, \mathbf{p}_{T(i)}) < 0.3 \tag{8}$$

where $\mathbf{E}_{T(i)} > 30$GeV and $i = 1, 2, 3$. This was employed as a check to make sure that any calculated $\not{H}_T$ was not simply a result of excluding a jet that sat just outside of the "good" jet cut definition ($E_T > 50$GeV, $F_{\text{EM}} < 0.9$). This is a good example of using a cut to protect against accepting events because of a self-imposed "real" missing $E_T$.

## 4 Extending the Search to an $n$-jet System

The application of $\alpha_T$ and $\Delta\phi_{(1,2)}$ to the dijet system, combined with the background reduction techniques discussed in [1], show the search channel to be promising for early SUSY searches at the CMS experiment. A natural extension to the channel comes from considering SUSY processes like

$$pp \to \tilde{g}\tilde{q} \to \tilde{q}q\tilde{N}q \to \tilde{N}qq\tilde{N}q \tag{9}$$

where the final state contains two LSPs and $n$ jets. This would obviously increase the potential signal yield which would be very useful for searches with early data where machine energy and luminosity may be below the design values. In order to do this, methods similar to those used in the dijet analysis for rejecting/constraining background processes must be established. This is the subject of the following section.

## 4.1 Extending the Preslection Cuts for $n$ Hadronic Jets

The following changes were made to the event preselection cuts:

- An event is classified as an $n$ jet system if it contains $n$ jets with $E_T > 50$GeV and $F_{\text{EM}} < 0.9$. Note that this makes any $n+1$ (i.e. third jet for the dijet system) cut redundant.
- Any event containing a jet with $E_T > 50$GeV and $F_{\text{EM}} > 0.9$ is rejected.
- The $\Delta\phi(\not{H}_T, \mathbf{p}_{T(i)})$ cut (Eq. 8) is applied to every jet up to and including the $n+1^{\text{th}}$ jet.

This defines our $n$-jet system. Note that all of the $n$ jets have the same cuts applied to them. The exception to this is the cut on the $\eta$ of the leading jet, $|\eta| < 2.5$.

## 4.2 The $n$-jet Philosophy

With the $n$ hadronic jets selected, the next step is to define a suitable discriminating variable upon which accept or reject events. We have taken the approach of combining the $n$ jets into a "pseudo-dijet" system and then applying the $\alpha_T$ and $\Delta\phi$ cuts to the pseudo-dijet. This is not an unreasonable approach to take, as all final state configurations will have emerged from an initial two-body interaction; the $n$ jets will, generally speaking, be the products of cascade/showering processes. Two questions arise:

1. How should a given number of jets be merged together to form a new "pseudo-jet"?
2. How should the particular combination of the $n$ jets to be merged into the two pseudo-jets be chosen?

We arrived at possible answers to these questions by considering the form of $\alpha_T$ as we look to apply it to the $n$-jet case.

## 4.3 Defining $\alpha_T$ for $n$ Jets

Firstly, we consider $M_{T(\text{Tot.})}$. It makes sense that this should be same for any combination of jets chosen for a given $n$-jet system. This condition is only met by employing the following merging scheme:

$$E_{T(jk)} = E_{T(j)} + E_{T(k)} \tag{10}$$

$$p_{x(jk)} = p_{x(j)} + p_{x(k)} \tag{11}$$

$$p_{y(jk)} = p_{y(j)} + p_{y(k)} \tag{12}$$

Figuratively speaking, this is the equivalent of adding together the length of the $\mathbf{p}_T$ vectors (which for massless objects, will have the same magnitude as the $E_T$) and pointing it in the direction of the vectorial combination of the $\mathbf{p}_T$ vectors. Note that we must now consider objects purely in the transverse plane, since $E_T$ depends on $\theta$. We therefore set $p_{z(jk)} = 0$, and so $E = E_T$ for the combined objects.

Likewise, the numerator of $\alpha_T$ suggests a strategy for choosing the most appropriate jet combination. We notice that the ideal QCD dijet case, where $\alpha_T = 0.5$, is obtained when $\Delta H_T = 0$ and the jets are balanced in $E_T$. So, we check all possible combinations of $n \to 2$ jets and select the one that produces the smallest $\Delta H_T$. Figure 3 shows how, for example, this picks out the most dijet-like combination for the 3 jet case. It should be noted that $\Delta H_T \neq 0$ for an $n$ jet system, as even small opening angles between jets that are merged will produce a larger (and so imbalanced) $E_{T(jk)}$. So the definition we arrive at for an $n$ jet $\alpha_T$ is

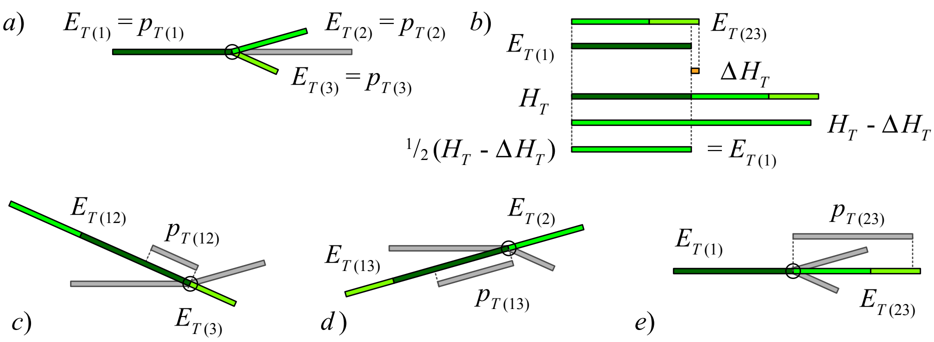


Figure 3: The $\Delta H_T$ jet clustering method illustrated. a) A QCD-like trijet event as viewed in the transverse plane; b) Calculation of the numerator of $\alpha_T$ using the minimum $\Delta H_T$, obtained from the most dijet-like combination $\{1,23\}$; c) The $\{3,12\}$ combination using the merging scheme of Equations 10, 11 and 12; d) The same for $\{2,12\}$; e) The same for $\{1,23\}$.

$$\alpha_{T(2,\ldots,n)} = \frac{\frac{1}{2}\left(H_T - \Delta H_{T(n)}\right)}{\sqrt{H_T^2 - |\not{H}_T|^2}} \tag{13}$$

where $\Delta H_{T(n)}$ is the minimum $\Delta H_T$ obtained by considering all possible $n \to 2$ jet combinations.

# 5 Preliminary Studies with CSA07 Monte Carlo Data

We have used the definition of $\alpha_T$ in Equation 13 in a preliminary analysis of the $n$ hadronic jet channel. The Monte Carlo datasets are the same as those used in [1], and the event preselection is essentially the same with the extensions listed in § 4.1. The results are shown in Table 1. Even without optimisation, it can be seen that

Table 1: Number of events (for $1\,\mathrm{fb}^{-1}$) passing successive cuts of $H_T > 500\mathrm{GeV}$ $\alpha_T > 0.55$, $0 < \Delta\phi < \frac{2\pi}{3}$ after the $n$ jet preselection cuts [1] have been applied.

| $n$ | Cut | QCD | $t\bar{t}, W, Z$ | $Z \to \nu\bar{\nu}$ | LM1 |
|---|---|---|---|---|---|
| 2 | $H_T$ | $3.3 \times 10^6$ | 245 | 2414 | 1770 |
| | $\alpha_T$ | 0 | 58.8 | 20.4 | 440.0 |
| | $\Delta\phi$ | 0 | 57.7 | 19.2 | 432.7 |
| 3 | $H_T$ | $6.8 \times 10^6$ | 213 | 5669 | 3071 |
| | $\alpha_T$ | 24.0 | 64.4 | 49.9 | 852.5 |
| | $\Delta\phi$ | 24.0 | 63.9 | 45.9 | 837.7 |
| 4 | $H_T$ | $4.0 \times 10^6$ | 86.0 | 7078 | 2510 |
| | $\alpha_T$ | 2.5 | 24.5 | 41.8 | 676.5 |
| | $\Delta\phi$ | 2.5 | 24.0 | 41.4 | 668.2 |
| 5 | $H_T$ | $1.0 \times 10^6$ | 19.2 | 4710 | 1350 |
| | $\alpha_T$ | 21.5 | 5.8 | 16.4 | 295.3 |
| | $\Delta\phi$ | 21.5 | 5.8 | 16.1 | 290.3 |
| 6 | $H_T$ | $1.8 \times 10^5$ | 2.6 | 2105 | 552.5 |
| | $\alpha_T$ | 0.4 | 0.8 | 8.4 | 103.1 |
| | $\Delta\phi$ | 0.4 | 0.8 | 8.2 | 101.0 |
| **Total** | $\alpha_T$ | 48.4 | 154.3 | 136.9 | 2367.4 |
| | $\Delta\phi$ | 48.4 | 152.2 | 130.8 | 2329.9 |

the approach looks promising, with $S/B = 7.0$ and $S/\sqrt{B} = 128$ after the $\Delta\phi$ cut. This is comparable to the dijet analysis, but with roughly five times the signal yield. Plots of $\alpha_T$ and $\Delta\phi$ for the $n = 4$ system are plotted in Figures 4 and 5 respectively. It is interesting to note that, as shown in Figure 5, even when the $\Delta H_T$ does not appear to reconstruct a back-to-back dijet system, the corresponding $\alpha_T$ value is low enough to make the event fail the cut. This point is made clearer by plotting $\alpha_T$ against $\Delta\phi$. Figure 6 shows this for the $n = 4$ system. So it would appear that when the method fails (for example, if a jet is missed, or a jet from the underlying event with

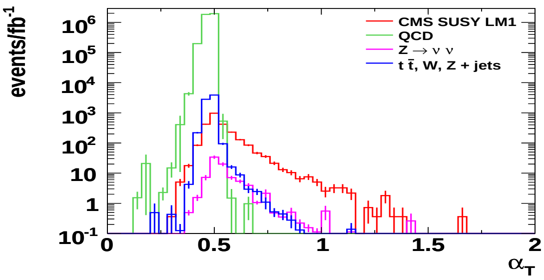


Figure 4: Plot of $\alpha_T$ for the $n = 4$ system.

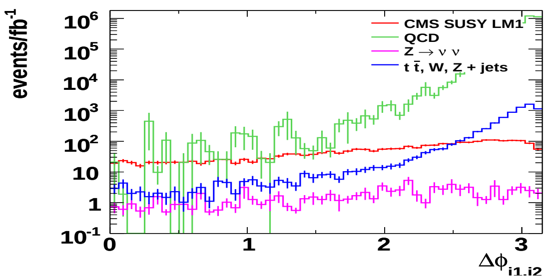


Figure 5: Plot of $\Delta\phi$ of the pseudo-dijets constructed with the min. $\Delta H_T$ method for the $n = 4$ system.

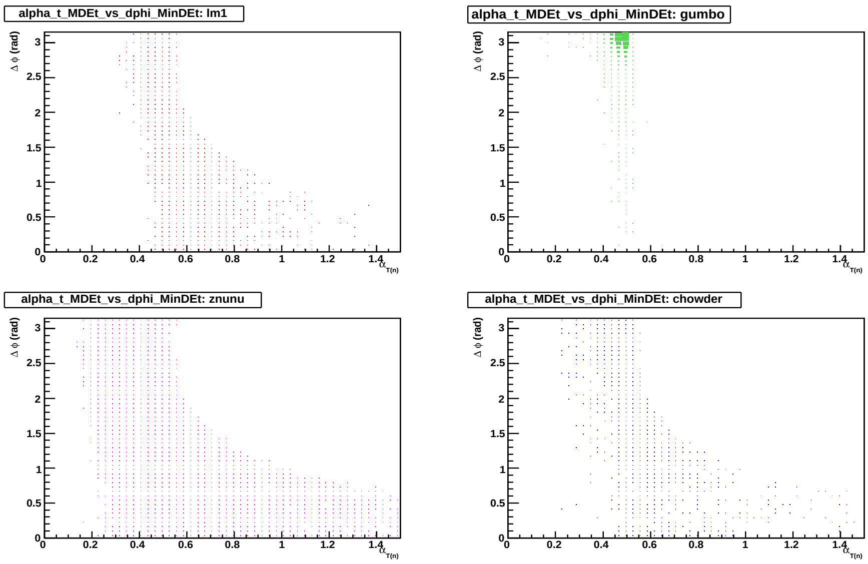


Figure 6: Plot of $\alpha_T$ against $\Delta\phi$ of the pseudo-dijet system as defined by the min. $\Delta H_T$ transverse clustering method for the signal and background MC samples.

$p_T > 50\mathrm{GeV}$ is included), by construction $\alpha_T$ rejects the event. This apparent "self-protection" is not unique to $\alpha_T$. The analysis was re-run without the $\Delta\phi(\not{H}_T, \mathbf{p}_{T(i)})$ cut, and considerably more background was found to be accepted (which was not generally the case for the dijet analysis). This is perhaps unsurprising; with the transverse energy of the interaction being split $n$ ways, jets just under the $E_T < 50\mathrm{GeV}$ selection cut are more likely to play an important part in contributing to fake $\not{H}_T$.

Observables such as $\Delta H_T$ and $\not{H}_T$, that display self-protection, could play a very useful role in early missing $M_T$ searches. A systematic study of these variables in the context of start-up conditions (jet smearing, detector noise, MC truth) would clearly be a worthwhile exercise.

# 6 Generalising the $\alpha_T$ Approach to Kinematics-based Searches

The form of $\alpha_T$ in Eq. 13 suggests that other combinations of kinematic variables may be worthy of investigation. We see that the linear combination of $H_T$ and $\Delta H_{T(n)}$ in the numerator, with $\Delta H_{T(n)} > 0$ for an $n$-jet system, will always favour lowering $\alpha_T$. This conservative approach will understandably lower the signal yield. To increase the signal yield, we can therefore imagine constructing the following $\alpha_T$-like observables,

$$\beta_T = \frac{\frac{1}{2}(H_T - \Delta H_T)}{H_T - |\not{H}_T|} \tag{14}$$

and

$$\zeta_T = \frac{\frac{1}{2}\sqrt{H_T^2 - \Delta H_T^2}}{\sqrt{H_T^2 - |\not{H}_T|^2}}. \tag{15}$$

We repeated the analysis described in for each new variable, cutting on $\beta_T > 0.8$ and $\zeta_T > 0.6$ respectively, and then $\Delta\phi$. These unoptimised values were chosen to reflect the anticipated kinematic "loosening" of the new observables. Preliminary results are shown in Table 2, and plots of $\beta_T$ and $\zeta_T$ for $n = 4$ are shown in Figures 7 and 8. The $\beta_T$ approach, after the (largely ineffectual) $\Delta\phi$ cut, has $S/B = 6.3$ and $S/\sqrt{B} = 145$ and, as expected, a 50% increase in the signal yield compared to $\alpha_T$. $\zeta_T$ produces very similar results to $\alpha_T$, but with a tighter cut value. This suggests $\zeta_T$ may be more robust; the peak around 0.5 for QCD events in $\zeta_T$ is certainly sharper, and likewise for $\beta_T$ the peak is wider. The robustness of each observable, and so the suitability/relevance for early searches, needs investigating further.

Table 2: Number of events (for $1\,\mathrm{fb}^{-1}$) passing successive cuts of $H_T > 500\mathrm{GeV}$, $\beta_T > 0.8$ or $\zeta_T > 0.6$ and $0 < \Delta\phi < \frac{2\pi}{3}$ after the $n$ jet preselection cuts [1] have been applied.

| | | $\beta_T > 0.8$ | | | | $\zeta_T > 0.6$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | Cut | QCD | $t\bar{t}, W, Z$ | $Z \to \nu\bar{\nu}$ | LM1 | QCD | $t\bar{t}, W, Z$ | $Z \to \nu\bar{\nu}$ | LM1 |
| 2 | $\beta_T/\zeta_T$ | 2.1 | 101.8 | 52.1 | 754.1 | 1.5 | 80.8 | 30.4 | 600.4 |
| | $\Delta\phi$ | 2.1 | 92.4 | 37.8 | 672.6 | 1.5 | 80.8 | 30.4 | 600.4 |
| 3 | $\beta_T/\zeta_T$ | 29.0 | 105.4 | 122.3 | 1339.3 | 6.0 | 69.8 | 47.9 | 916.3 |
| | $\Delta\phi$ | 27.5 | 88.4 | 82.4 | 1174.9 | 6.0 | 69.8 | 47.7 | 914.9 |
| 4 | $\beta_T/\zeta_T$ | 13.7 | 44.2 | 91.4 | 1068.5 | 2.5 | 21.1 | 26.9 | 556.4 |
| | $\Delta\phi$ | 7.7 | 37.0 | 76.6 | 940.5 | 1.0 | 21.1 | 26.9 | 555.0 |
| 5 | $\beta_T/\zeta_T$ | 24.0 | 7.9 | 38.0 | 462.7 | 21.5 | 4.0 | 7.9 | 176.0 |
| | $\Delta\phi$ | 22.0 | 7.5 | 28.9 | 408.6 | 21.0 | 4.0 | 7.9 | 176.0 |
| 6 | $\beta_T/\zeta_T$ | 2.5 | 0.9 | 16.2 | 151.5 | 0.4 | 0.3 | 2.8 | 46.5 |
| | $\Delta\phi$ | 2.5 | 0.9 | 13.6 | 138.1 | 0.4 | 0.3 | 2.8 | 46.5 |
| **Total** | $\beta_T/\zeta_T$ | 71.3 | 260.2 | 320.0 | 3776.1 | 31.9 | 176.0 | 115.9 | 2295.6 |
| | $\Delta\phi$ | 61.8 | 226.2 | 239.3 | 3334.7 | 29.9 | 176.0 | 115.7 | 2292.8 |

# 7 Conclusions and Outlook

At first glance, the approach taken to generalising the $\alpha_T$ observable of the hadronic dijet analysis to $n$-jets appears to offer a promising way of maintaining the signal-to-background ratio of the former while increasing the signal yield by a factor of five.

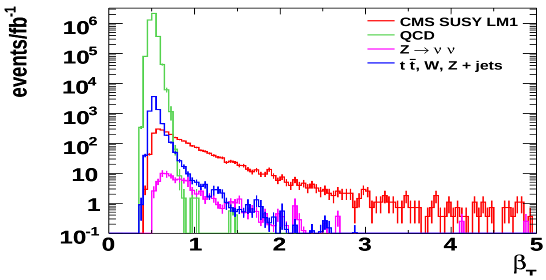


Figure 7: Plot of $\beta_T$ for the $n = 4$ system after all preselection cuts.

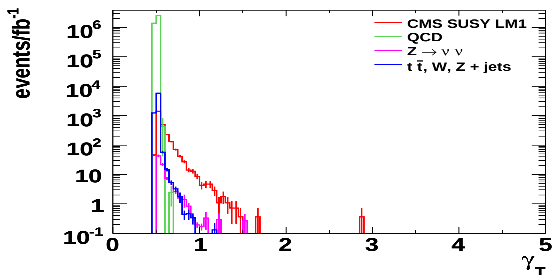


Figure 8: Plot of $\zeta_T$ for the $n = 4$ system after all preselection cuts.

The strategy also appears to have highlighted the potential of a class of “self-protecting” observables that mitigate for “fake” missing transverse energy caused by detector mismeasurements and noise, self-imposed cut choices, or incorrectly defined $n$-jet systems (i.e. the inclusion of underlying-event jets). However, much work remains to be done:

- The kinematics of the $\alpha_T$ variable, and the nature of the $\Delta H_T$ clustering method, needs to be understood - in particular, how $\alpha_T$ appears to self-correct even when an incorrect pseudo-dijet system is constructed.
- Other $\alpha_T$-like variables, like those suggested in § 6, need to be explored in terms of their utility in early SUSY searches.
- Methods for constraining the *real* $\not{H}_T$ backgrounds need to be established. A particularly interesting question is how the ABCD method, which relies on $\eta$ of the leading jet, may be extended to a pseudo-dijet system where $\eta$ cannot be defined. Initial studies have also shown that the $t\bar{t}$ SM background could be particularly troubling for the $n = 3, 4$ case.

These should be the subject of future studies.

# References

[1] H.Flaecher, J.Jones, T. Rommerskirchen, M.Stoye, *“SUSY Search with Dijet Events”*, **CMS AN-2008/071**.

[2] L.Randall, D.Tucker-Smith, *“Dijet Searches for Supersymmetry at the LHC”*, `arXiv:0806.1049`.

[3] The CMS Collaboration, G. L. Bayatian et al: *“CMS Technical Design Report, Vol. II: Physics Performance”*, J. Phys. **G34** (2007) 995-1579.