Available on CMS information server

CMS AN -2009/155

The Compact Muon Solenoid Experiment

Analysis Note

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

16 October 2009 (v2, 11 January 2010)

Study of the Correlation Between α_{T} and H_{T} Using Z+Jet Events

Paul Geffert and David Stuart

University of California, Santa Barbara

Abstract

We present a study of the correlation of α_{T}, a potentially very useful variable for rejecting QCD background in SUSY dijet searches, with H_{T} using Z+jet events. The Z+jet sample allows us to probe an H_{T} regime lower than is possible with dijet QCD events. We find that the α_{T} variable fails less often with increasing H_{T}, for the H_{T} region explored, and that this behavior could be confirmed with the first 100 pb^{-1} of data. We also find the behavior of α_{T} to be largely unchanged for different values of η, which would be helpful in certain methods of background estimation.

1 Introduction

In some regions of supersymmetry (SUSY) parameter space squarks can be pair produced and decay directly to a quark and lightest supersymmetric particle (LSP) [1], i.e. $\tilde{q} \rightarrow q + \tilde{\chi}^0$. Events of this type would produce a dijet signature due to the two quark jets as well as missing transverse energy ($\not{E}_{\mathrm{T}}$) due to the two LSP's. Because of its enormous cross section, dijet QCD is a major background to this signal.

α_{T} is a powerful variable for rejecting backgrounds, especially QCD, in SUSY dijet searches [2]. For dijet events, it is defined as

$$\alpha_{\mathrm{T}} = \frac{E_T^{j2}}{M_{T\mathrm{inv}}^{j1,j2}}, \tag{1}$$

where the two jets are ordered by descending E_{T} and $M_{T\mathrm{inv}}^{j1,j2}$ is the invariant mass of the two jets. For massless jets, we can rewrite this as

$$\alpha_{\mathrm{T}} = \frac{\sqrt{E_T^{j2}/E_T^{j1}}}{\sqrt{2(1-\cos\Delta\phi)}}, \tag{2}$$

where $\Delta\phi$ is the angle in phi between the two jets. α_{T} has been extended to include n-jet events, as described in [3].

In well measured QCD dijet events, the two jets should be back to back. Therefore, it can be seen from Eq. 2 that events of this type should have $\alpha_{\mathrm{T}} \leq 0.5$. However, due to jet energy and phi resolution effects, we will assume $\alpha_{\mathrm{T}} < 0.55$ for well measured QCD events from this point forward. In SUSY events, however, due to the high masses of the squarks, the resulting decay quarks are less likely to be back-to-back in phi than QCD dijets. From this knowledge and Eq. 2 it is apparent that $\alpha_{\mathrm{T}} > 0.55$ is possible for these SUSY events.

Despite the fact that $\alpha_{\mathrm{T}} < 0.55$ for well measured QCD dijet events, it is possible that poorly measured QCD events may be above the α_{T} threshold. Due to the E_T^{j2}/E_T^{j1} term in Eq. 2, α_{T} is somewhat insensitive to jet energy resolution effects. However, in the case that a third jet is lost or fluctuates below threshold, the $\Delta\phi$ between the remaining two jets will be less than π and could cause $\alpha_{\mathrm{T}} > 0.55$. If α_{T} is to be used as a background rejection tool, it will be important to understand how often it fails to reject background events. Because the SUSY signal would occur at high H_{T}, understanding how often α_{T} fails in QCD events as a function of H_{T} would be helpful.

This analysis will use Z+jet events to study the relationship between α_{T} and H_{T}. This could serve as an important cross check to a similar study using QCD dijet events performed in Ref. [2]. While our Z+jets data sample will have much lower statistics, it does have a number of advantages. For example, the Z's should be better measured than jets which allows for a higher degree of precision over the dijet study. This will allow for an exploration of an H_{T} regime lower than what is reasonable for the dijet study. It is also important that no new physics is expected in Z+jet events for this low H_{T} regime. Finally, the precision of the Z's should allow for exploration of α_{T} failures in the data.

This note is organized as follows: Section 2 describes the event sample and the details of event selection. Sections 3 and 4 discuss the correlation of α_{T} with H_{T} and $|\eta|$, respectively. A simple model relating the likelihood of α_{T} failing to H_{T} can be found in Sec. 5. The issue of jet energy scaling is briefly discussed in Sec. 6 before concluding in Sec. 7.

2 Event Selection

The Monte Carlo (MC) sample used in this analysis is 800 pb^{-1} of $Z \rightarrow \mu\mu$ events at $\sqrt{s} =$ 10 TeV [4]. This sample was generated using PYTHIA, the detector response simulated with FullSim, and the result ntupled with CMSSW 2.2.6.

The selection criteria for our Z+1jet events are:

- Exactly two muons passing loose muon requirements that reconstruct to a Z in the mass window [81,101] GeV. The loose muon requirements are:
  - $p_T > 10$ GeV
  - $|\eta| < 2.4$
  - $d_0 < 0.2$ cm
  - $\chi^2/ndof < 10$
  - $NumValHits > 10$
  - $GlobalMuonPromptTight \neq 0$
- The event is vetoed if there are any electrons passing the loose requirements. The loose electron requirements are:
  - $p_T > 10$ GeV
  - $|\eta| < 2.5$
  - $d_0 < 0.2$ cm
  - $robustLooseId \neq 0$
- Z $p_{\mathrm{T}} > 30$ GeV
- Exactly one jet with $p_{\mathrm{T}} > 30$ GeV

To help eliminate backgrounds, a sideband subtraction is performed using events that pass all of the above requirements with the alteration that the reconstructed Z mass be in one of the [71,81] GeV or [101,111] GeV windows. Events that fall into one of these sidebands will be included in distributions with the usual Z mass events, but with a weight of -1. This will help reduce the effects of backgrounds such as $t\bar{t}$, WW, or WZ where a Z lepton is lost. These types of events have real $\not{E}_{\mathrm{T}}$ and hence can easily have $\alpha_{\mathrm{T}} > 0.55$. This is not of particular importance in the Monte Carlo but could have a significant effect on the data.

Because we are using Z+1jet events to mimic dijet events, it is necessary to slightly alter the definitions of a couple of kinematic variables. For instance, we define $H_{\mathrm{T}} = Zp_T + p_T^{\mathrm{jet}}$ as if the Z p_{T} was from a jet. Also, we do not use the definition of α_{T} in Eq. 1 because we want to use the massive Z to mimic a massless jet. Instead we use Eq. 2 where the Z p_{T} is substituted for the appropriate jet p_{T}.

3 Correlation of α_{T} and H_{T}

As mentioned in the introduction, we expect SUSY events to have high H_{T}. This makes it particularly important to understand the behavior of the α_{T} variable as a function of H_{T}. In other words, if a dijet search using α_{T} finds an excess of events at high α_{T} and H_{T}, we would like to have confidence that this excess comes from new physics and not our QCD background. As an example of the correlation between H_{T} and α_{T}, Fig. 1 shows the α_{T} distributions after all event selection requirements for two different regions of H_{T}. The two distributions are significantly different with the higher H_{T} distribution having a smaller tail at high α_{T}.

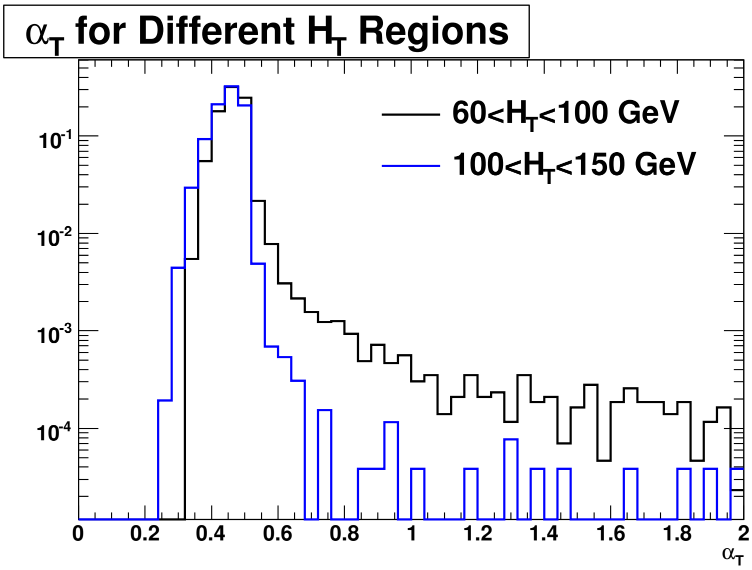


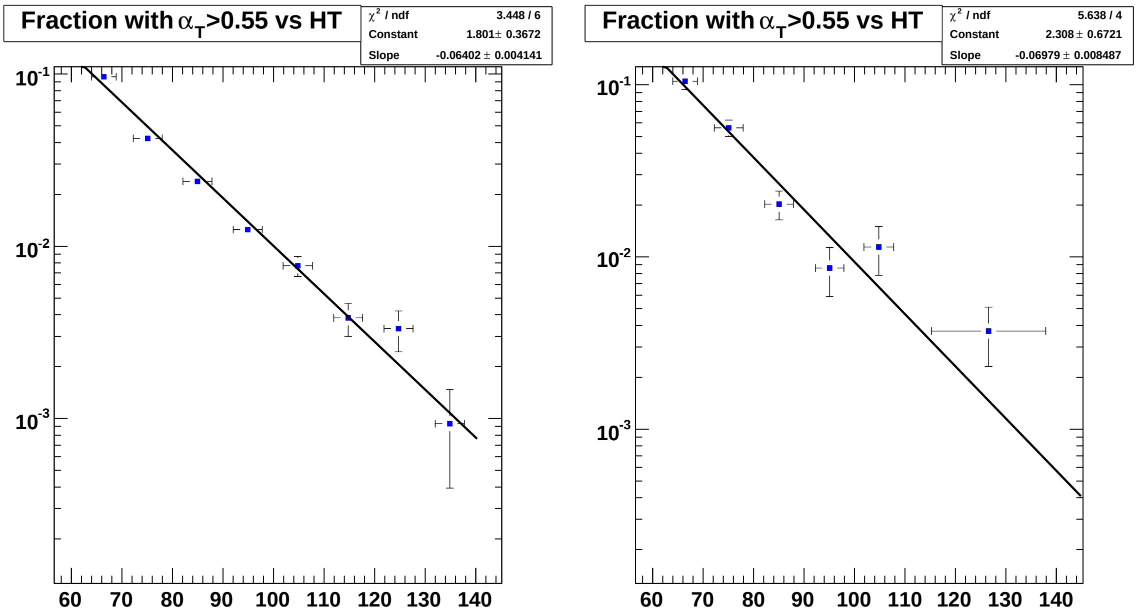
Figure 1: The α_{T} distributions for two different regions of H_{T} after all event selection requirements. Both distributions are normalized to unity for ease of comparison. We can see that the two distributions are significantly different with the higher H_{T} distribution having a smaller fraction of events with $\alpha_{\mathrm{T}} > 0.55$.

To help quantify the correlation of α_{T} and H_{T}, we introduce a function $f(H_{\mathrm{T}})$ that represents the fraction of events with $\alpha_{\mathrm{T}} > 0.55$ at a given H_{T}. By keeping track of the number of events with α_{T} above and below 0.55 for different H_{T} bins, we can use our sample to determine the behavior of $f(H_{\mathrm{T}})$. The results of this for the full 800 pb^{-1} sample are shown in Figure 2(a). For the H_{T} regime shown, $f(H_{\mathrm{T}})$ is a decreasing function that is well approximated by an exponential. A simple model that may help account for this behavior is described in Sec. 5.

Figure 2(b) shows $f(H_{\mathrm{T}})$ vs H_{T} for only 100 pb^{-1}. Even with these limited statistics, it is still possible to see that $f(H_{\mathrm{T}})$ is a decreasing function somewhat consistent with an exponential up to $H_T \sim 120$ GeV. This figure gives us hope that we might be able to see this relationship with early data.

4 Correlation of α_{T} and η

The α_{T} based dijet analysis in Ref. [2] uses $f(H_{\mathrm{T}})$ as a function of lead jet $|\eta|$ to normalize the background estimation. For their method, it is important that the distribution be relatively flat in $|\eta|$ for the backgrounds. As a cross check, we would like to use our Z+jet sample to test for flatness in $|\eta|$. Our H_{T} regime will be much lower than that used in the SUSY analysis, but we should not have any contamination from new physics. Also, this check can be preformed with real data once it is available. We introduce the function $f(|\eta|, H_{\mathrm{T}})$ which represents the fraction of events with $\alpha_{\mathrm{T}} > 0.55$ at a given $|\eta|$ and H_{T}. In this analysis, we take $|\eta|$ from the Z or jet, whichever has higher p_{T}. Figure 3(a) shows $f(|\eta|, H_{\mathrm{T}})$ for three different H_{T} regimes for the full dataset. In all cases, the result seems consistent with flatness in $|\eta|$. Figure 3(b) shows $f(|\eta|, H_{\mathrm{T}})$ for only 100 pb^{-1}. Because of the limited statistics, it is difficult to tell if $f(|\eta|, H_{\mathrm{T}})$ is flat for any of the H_{T} regimes shown.



(a) (b)

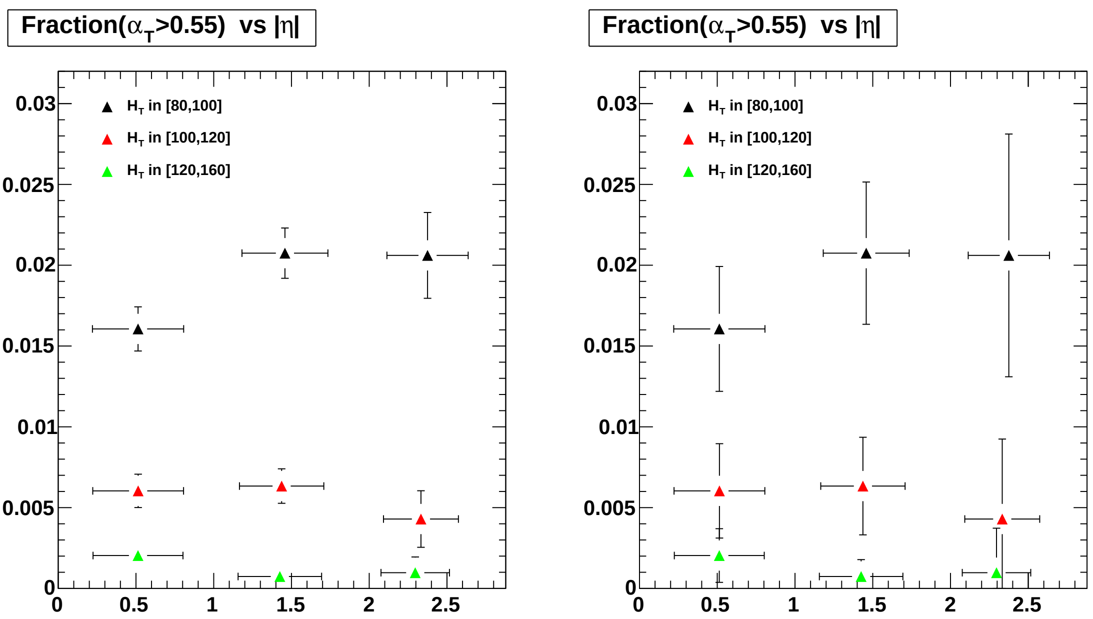
Figure 2: $f(H_T)$ vs H_T for (a) the entire 800 pb^{-1} sample and (b) 100 pb^{-1}. We can clearly see that $f(H_T)$ is a decreasing function and is consistent with an exponential in the H_T regime shown. Values of H_T below 60 GeV are excluded by the selection cuts. The H_T region above 140 GeV is statistically imprecise and not shown.

5 Simple Model for Behavior of $f(H_T)$

5.1 Lost Jets

As mentioned previously, one scenario that can lead to an event failing the α_T cut occurs when there is an additional jet in the event that gets lost. In our Z+jets sample, this would occur when a Z+2jet event loses a jet and the remaining jet and the Z are sufficiently close in phi so that α_T falls above 0.55. In an event with a lost jet that is otherwise well measured, the remaining jet and Z must balance the lost jet in the transverse plane. An increase in p_T of the lost jet decreases the phi angle between the remaining jet and Z, which in turn increases α_T. In order for an event in our sample to fail α_T due to a lost jet, the p_T of the lost jet must be great enough to reduce the phi angle between the Z and remaining jet enough so that α_T $>$ 0.55. The p_T required to do this scales linearly with the H_T of the event. Therefore, as H_T increases, the likelihood of losing such a jet is highly correlated with the jet p_T spectrum, which is similar to an exponential.

To study these types of events, we can utilize a Z+2jet sample and remove jets to obtain a Z+1jet sample where each event has a missing jet. Therefore, events failing α_T in this sample will be due to the loss of a jet with overwhelming probability. Figure 4 was produced by removing jets from a Z+2jet sample in two different ways, each with linear probability in p_T. Because exponentials dominate lines, the exponential nature of the jet p_T spectrum should be the determining factor that influences $f(H_T)$. The sample with red points has 50% of 30 GeV jets removed with the removal probability linearly decreasing to 0% of 200 GeV jets removed.



(a) (b)

Figure 3: $f(|\eta|, H_T)$ vs $|\eta|$ for (a) the 800 pb^{-1} sample and (b) 100 pb^{-1}. We can see that for the full sample $f(|\eta|, H_T)$ is consistent with flatness in $|\eta|$ for each of the H_T regimes shown. However for 100 pb^{-1} the extremely large errors make it difficult to tell if $f(|\eta|, H_T)$ is consistent with flatness in $|\eta|$. Note that the central values of $f(|\eta|, H_T)$ in (b) are taken from (a) with the errors scaled up by a factor of $\sqrt{8}$ to reflect the change in sample size.

The blue points have 100% of 30 GeV jets removed to 0% of 200 GeV jets removed. In both cases, the result is a distribution of $f(H_T)$ that is consistent with an exponential.

One complication of this simple model is that the removed jet p_T spectrum will be different for different regions of H_T. As an example, we should have that the removed jet p_T is less than the H_T of the Z and remaining jet due to momentum conservation. This effect of a variable removed jet p_T spectrum can potentially create additional functional forms for $f(H_T)$, possibly to the extent that $f(H_T)$ is not reliably exponential. It may be possible to study this by varying the way jets are lost in the Z+2jet sample.

5.2 Soft Jets

Another way that QCD and Z+jet events can have $\alpha_T > 0.55$ is multiple soft jets aligning with similar phi angles. The Z and jet recoil against these soft jets in the transverse plane to give essentially the same effect as losing a single hard jet, as described above. Soft jets will have low p_T and hence will not pass our jet requirements, but will still deposit energy in the calorimeter. This means that the $\not{E}_T$ should account, to some extent, for the soft jets while the $\not{H}_T$ will not. (In Z+1jet events, $\not{H}_T = -(\vec{p_T}^Z + \vec{p_T}^{\mathrm{jet}})$.) Therefore, events with $\alpha_T > 0.55$ where the $\not{H}_T$ and $\not{E}_T$ are largely uncorrelated in magnitude and/or phi may have failed due to soft jets.

To study these types of events, we create a soft jet enhanced sample by adding soft jets to

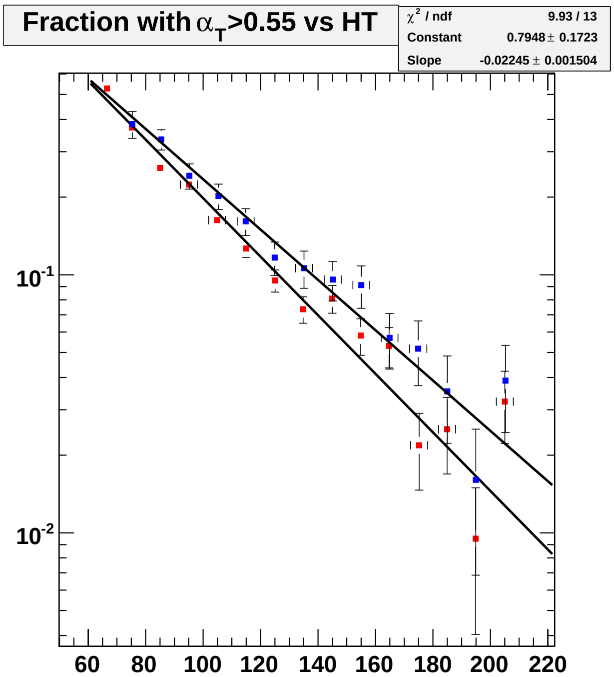


Figure 4: $f(H_T)$ vs H_T for Z+2jet events where one of the jets has been removed. The sample with red points has 50% of 30 GeV jets removed with the removal probability linearly decreasing to 0% of 200 GeV jets removed. The blue points have 100% of 30 GeV jets removed to 0% of 200 GeV jets removed. In both cases, the result is consistent with an exponential.

our usual Z+1jet sample. There are potentially multiple ways in which this can be done, but we have chosen the following method. Let us assume, for the remainder of this paragraph, that a soft jet with some particular p_T and phi has already been chosen. We first determine the invariant transverse mass of the Z, jet, and $\not{H}_T$. Then we create a 4 vector where this invariant mass is given equal and opposite momentum to the soft jet. The resulting boost of this vector is given to the Z and jet and the H_T and $\not{H}_T$ are recomputed to finish the process. This procedure can be applied iteratively to add as many soft jets as desired.

One detail of the above algorithm is whether or not to include the Z mass in the calculation of the invariant transverse mass and the final boosting of the Z, as described in the previous paragraph. Because we are interested in mimicking the effects of soft jets on dijet QCD events, in which the jets are taken to posses negligible mass, it makes some sense to ignore the Z mass in the above procedure. However, as a cross check, we have tested the method with and without including the Z mass and found the results to be comparable. Note that all plots related to soft jets in this note have been produced ignoring the Z mass.

Using the above method, we can vary the p_T distribution and number of soft jets that are added to events and observe the effects on $f(H_T)$. Because the jet p_T threshold is 30 GeV, we require that a soft jet added to an event have $p_T < 30$ GeV. In all cases, the soft jets will have a distribution random in phi. Figure 5 shows $f(H_T)$ vs H_T for two soft jet enriched samples with different distributions of the number of soft jets and soft jet p_T. The sample with red points has an exponential distribution with mean 4 for the number of soft jets and a soft jet p_T distribution that is exponential with mean 30. The sample with blue points has an exponential number of soft jets distribution with mean 3 and a soft jet p_T distribution that

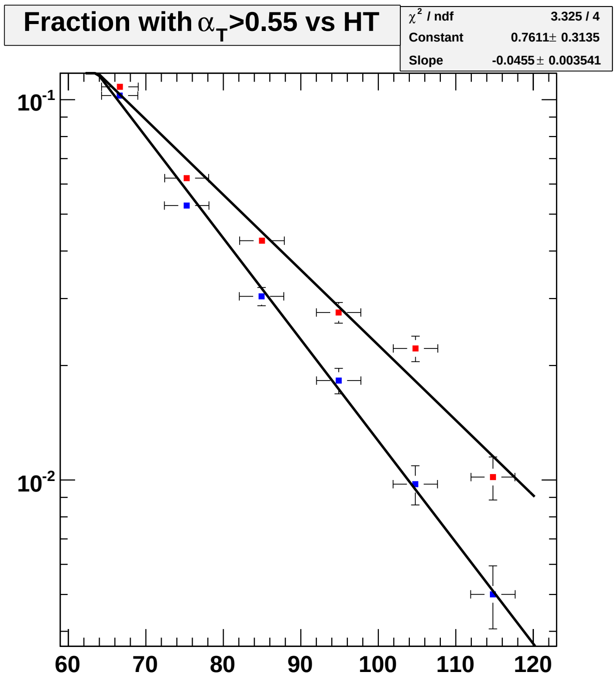


Figure 5: $f(H_T)$ vs H_T for Z+1jet events where soft jets have been added. The sample with red points has an exponential distribution with mean 4 for the number of soft jets and a soft jet p_T distribution that is exponential with mean 30. The sample with blue points has an exponential number of soft jets distribution with mean 3 and a soft jet p_T distribution that is exponential with mean 10. Note that added soft jets must have $p_T < 30$ GeV in order to stay below threshold. In both cases, the result is consistent with an exponential.

is exponential with mean 10. Both of these samples have $f(H_T)$ consistent with a decreasing exponential. The decrease of $f(H_T)$ as H_T increases makes sense because events with higher H_T are affected less by adding a soft jet.

5.3 Combined Effects of Lost and Soft Jets

In our Z+1jet sample there will be events that fail due to a combination of losing a jet and soft jets. Therefore it is important to study the behavior of $f(H_T)$ in a sample that is enriched with both lost jets and soft jets. For completeness, different combinations of jet removal and soft jet addition should be considered. As an example, Fig. 6 shows $f(H_T)$ vs H_T for two such combinations of jet removal and soft jet addition. To produce this plot, we use a Z+2jet sample and first remove jets and then add soft jets to the resulting Z+1jet events. Both the red and blue sets of points are consistent with an exponential decrease in $f(H_T)$. This result is encouraging because it shows that the samples enriched simultaneously in both lost jets and soft jets that we created still exhibit $f(H_T)$ that decreases with increasing H_T. This helps to demonstrate the robustness of α_T.

6 Jet Energy Scale

One important effect when this study is performed with actual data will be the jet energy scale. The reason for this is that, unlike dijets, Z+jets does not use only the hadronic calorime-

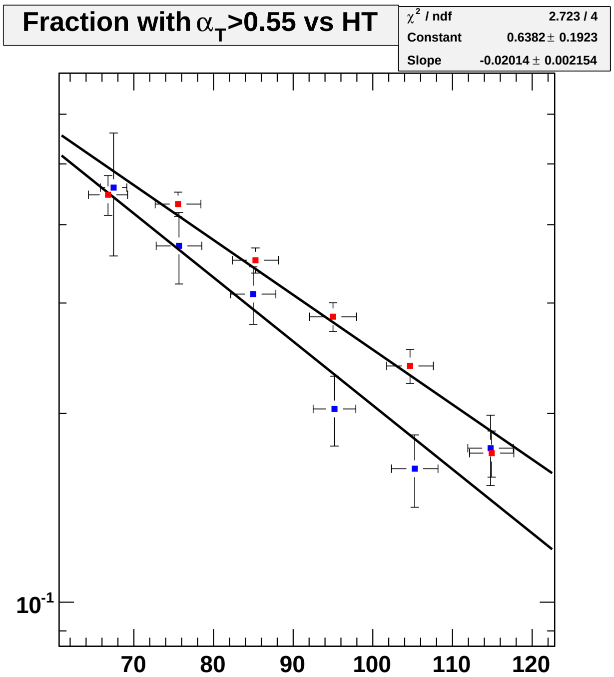


Figure 6: $f(H_\mathrm{T})$ vs H_T for events enriched in both soft jets and lost jets. We start with a sample of Z+2jet events and remove jets before soft jets are added to the resulting Z+1jet events. The sample with red points has 50% of 30 GeV jets removed with the removal probability linearly decreasing to 0% of 200 GeV jets removed and then an exponential distribution with mean 4 for the number of soft jets added with soft jet p_T distribution that is exponential with mean 30. The sample with blue points has 100% of 30 GeV jets removed to 0% of 200 GeV jets removed and then an exponential distribution with mean 3 for the number of soft jets added with soft jet p_T distribution that is exponential with mean 10. In both cases, the result is consistent with an exponential.

ter for energy measurements. This effect will be particularly important in early data when the jet energy scale has greater uncertainty.

Figure 7 shows the Z+1jet sample α_T distributions for different jet energy scales. The distributions are normalized to the sample with no jet energy scaling for ease of comparison. There are no drastic differences between the three distributions. The two scaled distributions both have an excess over the unscaled sample at $\alpha_\mathrm{T} < 0.4$. This shift to lower α_T of the two scaled samples makes sense in terms of Eq. 2. With the scaling of jet energies, the E_T ratio in Eq. 2 should decrease for events that are well balanced before the scaling.

The effect of jet energy scaling on $f(H_\mathrm{T})$ is shown in Fig. 8, which was produced using the same jet energy scaled samples as Fig. 7. For each sample, $f(H_\mathrm{T})$ clearly behaves as a decreasing function in the H_T range shown. Also, each sample is consistent with an exponential. In light of Fig. 8, it seems that a reasonable jet energy scale offset should not drastically hinder our study.

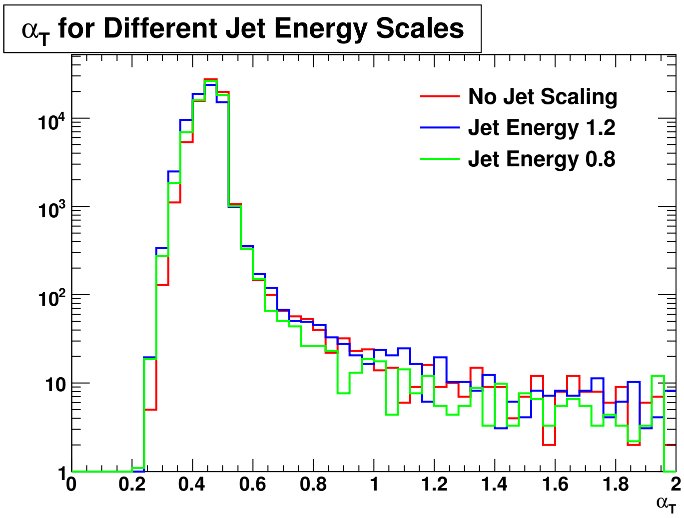


Figure 7: The α_T distributions for various levels of jet energy scaling. There are no drastic differences between the three distributions. The two scaled distributions both have an excess over the unscaled sample at $\alpha_T < 0.4$, which we believe we understand.

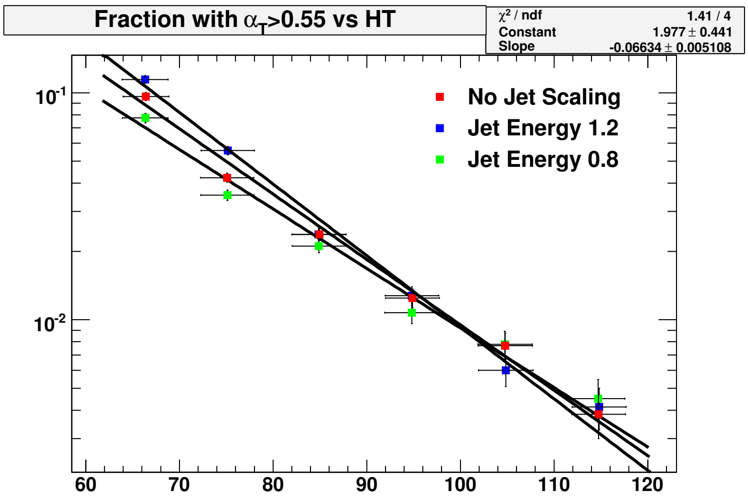


Figure 8: $f(H_T)$ vs H_T for various levels of jet energy scaling. All three samples have $f(H_T)$ clearly decreasing and consistent with an exponential.

7 Conclusions

We have presented a study of the relationship between failures of α_T and H_T as well as $|\eta|$ for Z+jet events. We found that $f(H_T)$ decreases in a way consistent with an exponential for the H_T regime studied and that this could potentially be seen with 100 pb^{-1} of data. Also, $f(|\eta|, H_T)$ was found to be relatively flat in $|\eta|$ for fixed H_T in the observed H_T regions. However, with only 100 pb^{-1} of data, it will be difficult to determine such.

We have developed methods to intentionally enrich events with lost jets and/or soft jets so their effects on $f(H_T)$ can be studied. In doing so, we have seen $f(H_T)$ distributions consistent with falling exponentials for different combinations of these two effects. These methods can be applied to enrich actual data, when it is available, with lost jets and soft jets as an important check of the behavior of $f(H_{\mathrm{T}})$.

This study could be an important cross check to the analogous dijet study once data is available. In the future, it may be possible to use γ+jet events as an additional handle on α_{T} in complement to the Z+jet and dijet studies.

References

[1] L. Randall and D. Tucker-Smith, "Dijet Searches for Supersymmetry at the LHC," *Phys. Rev. Lett.* **101** (2008) 221803, arXiv:0806.1049.

[2] CMS Analysis Note, CMS-AN 2008/071.

[3] CMS Analysis Note, CMS-AN 2008/114.

[4] We use Zmumu_Summer08_IDEAL_V9_v1/GEN-SIM-RECO.