VariableR Reclustering in Multiple Top Quark and W Boson Events

Jeremy Hyde

August 14, 2015

Abstract

VariableR jet reclustering is an innovative technique that allows for the reconstruction of boosted object over a wide range of kinematic regimes. Such capability enables the efficient identification of events with multiple boosted top quarks which is a typical signature for new physics processes such as the production of the supersymmetric partner of the gluon. In order to evaluate the performance of the algorithm, the VariableR reclustered jets are compared with fixed radius reclustered jets. The flexibility of the algorithm is tested by reconstructing both boosted top quarks and boosted W bosons. The VariableR reclustering method is found to be more efficient than the fixed radius algorithm at identifying top quarks and W bosons in events with four top quarks, therefore enhancing the sensitivity for gluino searches.

1 Introduction

Supersymmetry is a possible extension to the Standard Model (SM), which hopes to provide a solution to some of the underlying problems in the current model. One of the most promising discovery channels for the supersymmetric partners posited by supersymmetry is the strong production of gluinos that decay into top quarks and neutralinos.

Figure 1: Feynman diagrams for gluino pair production with intermediate virtual top squarks and a final state of four boosted top quarks.

SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program, under Contract No. DE-AC02-76SF00515.

The final state for the process includes up to four boosted top quarks depending on the gluino and neutralino masses. Efficient reconstruction of multiple boosted objects is crucial for searches based on this signature.

2 Jets

One of the basic problems in particle physics is using the outgoing long lived particles to reconstruct the shorter lived intermediate particles produced by a collision. In the case shown in Fig. 1 the process under study produces quarks, however, due to color confinement and interactions with the strong nuclear force, these quarks cannot exist individually and instead will form pairs or triplets of particles. This causes a shower of hadrons which interact with the detector. This stream of hadrons is called a jet.

Therefore, instead of a single impact on the detector from a quark, multiple energy deposits from the hadrons are observed. These energy deposits are reclustered into 0.4 radius jets in the η - ϕ space using the anti- k_T algorithm[2] based on the distance metric defined below.

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (1)$$

The decay of an object with high transverse momentum (p_T), referred to as boosted objects, produces decay products in a narrow cone. The typical distance between these decay products, the cone size (ΔR) depends on the mass m and p_T of the object with the follow relationship

$$\Delta R \sim \frac{\rho}{p_T} = \frac{2M}{p_T}, \quad (2)$$

where ρ represents the typical scale of the boosted object decay. In the case of a top quark, which decays into a W boson and bottom quark, there will typically be 3 jets (two from the W boson decay and one from the bottom quark decay).

3 Reclustering

Jet reclustering[4] consists of reconstructing larger jets by clustering calibrated $R = 0.4$ jets. As shown in Fig. 2 from Ref.[4] the energy deposits are reclustered into jets of $R = 0.4$, these jets are then used as inputs to form the larger jets of $R = 1.0$. Such reclustering of jets are referred to in this report as fixed radius reclustering.

Figure 2: An example from Ref.[4] event which has been clustered using the anti- k_T $R = 0.3$. The shaded regions show the jet area determined by clustering particles. The stars shows the anti- k_T reclustering jet of $R = 1.0$.

3.1 VariableR

VariableR reclustering[3] is able to reconstruct a wide kinematic regime due to the range of final reclustering radii it can produce. By varying the mass value used in the radius calculation seen in Eq. 2 the decay of specific particles can be targeted. For reconstructing top quarks, the $\rho = 2 \times 173.5$ GeV is used to define the characteristic radius of the large R jet. For reconstructing W bosons ρ is set to 2×80.4 GeV. The range of possible radii in VariableR reclustering is limited at the lower end by the size of the input $R = 0.4$ jets, while an upper limit of 1.5 is applied corresponding to the minimum p_T requirement.

4 Samples

The sample used as signal is simulated with the MadGraph[1] generator. It consists of events where a gluino pair. Each gluino decays into two tops and a neutralino (see Fig. 1). The background, the SM production of top quark pairs, is simulated with Powheg[5]. In order to ensure sufficient statistics in the signal region two background samples were created. The signal and background event rates are normalized to the respective cross-sections (σ taking into account the generated filter efficiency (ϵ) and assuming an integrated luminosity ($\mathcal{L}$) of 5 fb^{-1} .

$$SF = \frac{\mathcal{L} * \epsilon * \sigma}{n_{generated}} \quad (3)$$

5 Top Quark Reconstruction

5.1 Mass Distribution

In order to understand how VariableR performs with respect to fixed reclustering, the mass distributions of the various top quark candidates are compared. Figure 3 shows the mass distribution of the four highest p_T top quark candidates in each event. Both fixed and VariableR reclustering produce sharp peaks around the top quark mass, which is a good indicator that these candidates are indeed top quarks. The peak is particularly prominent for the two highest p_T candidates. However, the fixed reclustering candidates with the highest p_T in each event often has a much larger mass than the VariableR reclustering ones. Instead, in the fourth highest p_T candidate, very few fixed reclustering candidates are around the top quark mass while the VariableR candidates still cluster around the top quark mass.

Figure 3: The mass distributions of the four highest p_T reclustered jets in signal events (in order from top left to bottom right). The distributions for VariableR reclustered jets are shown in red and the fixed reclustered jets one is shown in blue. A minimum p_T of 300 GeV is required for the two candidates with highest p_T and a minimum p_T of 200 GeV is required for the third and fourth highest p_T candidate.

The large mass values observed for the highest p_T fixed reclustering candidates are due to accidental reclustering of jets from different top quarks into one object.

Figure 4: An event display with $p = 2 \times \text{Top quark mass}$ where the black circles are the fixed reclustered jets and the black triangles represent their constituents. The red circles represent the VariableR reclustered jets and the red crosses their constituents. The numbers at the masses associated with their corresponding reclustered jets.

Fig. 4 is an example of an event where this occurs. The mass of the fixed radius reclustered object is 429 GeV, which is much higher than the top quark mass we expect, and a fourth fixed radius candidate is not reconstructed. In other cases the fourth candidate is reconstructed, but with a lower mass.

On the other hand the VariableR reclustering algorithm finds two candidates in the same region with masses of 192 GeV and 198 GeV because VariableR uses a range of radius, which allows it to better resolve nearby jets.

There is also a peak in the third and fourth highest p_T mass distributions at very low mass which is most likely due to the reclustering of random jets. Therefore in the following analysis, top quark candidates are required to have a mass of at least 50 GeV

5.2 Reconstruction Efficiency

The algorithm's performance is quantified as the efficiency of reconstructing multiple top quark events. Fig. 5 shows the reconstruction efficiency for both the VariableR and fixed reclustering methods.

The range of radii VariableR uses allows it to recluster jets more efficiently than fixed radius reclustering. The requirement of four or more top quark candidates only allows fixed reclustering to select 10% of the signal events while VariableR reclustering is able to select around 30% of the signal events.

Figure 5: Reconstruction efficiency of top quark reclustering with $m > 50$ GeV

5.3 Significance

In order to investigate the application of the VariableR algorithm for signal discrimination, the signal significance is calculated as

$$Significance = \frac{signal}{\sqrt{signal + background}} \quad (4)$$

Figure 6: The signal significance for top quark reconstruction with VariableR reclustering shown in a solid red line and fixed reclustering shown in a dotted red line

Fig. 6 shows that fixed and VariableR reclustering methods produce relatively similar significance for requirement of one or two top quark candidates, however when requiring three or four top quark candidates the significance when using the VariableR algorithm is higher. The fixed reclustering produces less reclustered objects on average as it is reclustering multiple top quarks together.

6 W Boson Reconstruction

6.1 Mass Comparison

In this section, the capability to identify W bosons of the VariableR and fixed reclustering algorithms is studied. Fig. 7 shows the mass distribution of the highest p_T W boson candidate for each event. The fixed reclustering mass distribution remains the same as in Fig. 3 as it is not dependent on ρ , however, the VariableR now has a peak centered at approximately 80 GeV which is the mass of a W boson.

Figure 7: The mass distribution of the highest p_T reclustered jet for W boson reconstruction. The VariableR reclustering is shown in red while the fixed reclustering is shown in blue. A minimum p_T requirement of 200 GeV is also placed on the candidates.

Fig. 8 displays an example of W boson reconstruction. The fixed reclustering algorithm identifies three objects: one of top quark mass, one of W boson mass and one of mass 52 GeV. However, VariableR reclustering algorithm identifies 2 objects of W boson mass and the other 52 GeV object. Not only was VariableR able to recluster the same W boson as the fixed reclustering but it is also able to recluster the W boson that is a decay product of the fixed reclustering identified top quark. This is an example of VariableR reclustering being able to reconstruct W bosons in various regimes.

6.2 Reconstruction Efficiency

The next step is to compare how efficiently fixed and VariableR reclustering are able to reconstruct W bosons. In addition to the previously discussed minimum requirement on mass of 50 GeV, an upper bound was added of 110 GeV in order to target W bosons. Fig. 9 that for the requirement of four or more W boson candidates 35% of the signal events are selected using VariableR reclustering

Figure 8: An event display with $\rho = 2 \times W$ boson mass where the black circles are the fixed reclustered jets and the black triangles represent their constituents. The red circles represent the VariableR reclustered jets and the red crosses their constituents. The numbers at the masses associated with their corresponding reclustered jets.

but only 0.5% of the signal events are selected for fixed reclustering. VariableR is targeting W bosons while the fixed radius reclustering depends on the p_T range of boosted object.

Figure 9: Reconstruction efficiency of W boson reclustering with $m > 50$ GeV

6.3 Significance

The difference seen in reconstruction efficiency translates into the signal significance comparison shown in Fig. 10. The significance increases with the minimum number of required reclustered jets for VariableR reclustering, but decreases for fixed radius reclustering. When a minimum of four reclustered jets

is required the VariableR signal significance is approximately 15 times higher than the fixed reclustering signal significance.

Figure 10: The signal significance for W boson reconstruction with VariableR reclustering shown in a solid blue line and fixed reclustering shown in a dotted blue line

7 Comparison

Fig. 11 compares the signal of the W boson reconstruction and the signal of the top quark reconstruction for both fixed radius and VariableR reclustering. For any minimum number of reclustered jets, the top quark identification performs better than the W boson identification. This is due to the decrease in efficiency caused by the mass window applied in W boson reconstruction.

Figure 11: The signal significance comparison for W boson and top quark reconstruction

Furthermore the signal of the top quark fixed reclustering and the signal of the W boson VariableR reclustering are approximately equal when a minimum of four reclustered jets are required.

8 Acknowledgements

- Francesco Rubbo
- Ariel Schwartzman
- Su Dong
- Ben Nachman
- SLAC, SLUI, DOE and CERN

References

- [1] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [2] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.
- [3] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with Variable R. *JHEP*, 06:059, 2009.
- [4] Benjamin Nachman, Pascal Nef, Ariel Schwartzman, Maximilian Swiatkowski, and Chaowaroj Wanotayaroj. Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC. *JHEP*, 02:075, 2015.
- [5] Carlo Oleari. The POWHEG-BOX. *Nucl. Phys. Proc. Suppl.*, 205-206:36–41, 2010.