Boosted Objects in DESY-CMS-SUSY Searches in a Single Lepton Final State

Sophie Williamson - Summer Student Project

DESY, Notkestraße 85, 22607, Hamburg, Germany
Department of Physics and Astronomy, University of Sussex, BN1 9QH, Brighton, UK.

ABSTRACT: This project aims to investigate boosted objects, such as top quarks and W bosons, in current searches for supersymmetry signals with the CMS detector at the LHC. Using Monte Carlo simulations, we check out expected correlations between generator level quantities such as mass, transverse momentum and the distance between decay products. In this project we focus on a simplified SUSY model, where pair produced gluinos decay into a neutrino and two top quarks and compare this to the main background originating from ttbar and jet events. After checking generator level quantities we look at reconstructed jets and match them to MC truth tops and W's for different variables. The main conclusion of this project is to suggest that investigations should be carried out into using a smaller cone size for W jets than the default size used by the CMS collaboration.

Contents

1 Motivation for Supersymmetry

1.1 The Standard Model

The Standard Model (SM) of particle physics is the best model we have to date classifying all known subatomic particles as well as describing the strong, weak and electromagnetic (EM) interactions. To this point, the SM has been self-consistent and in 2012, CMS and ATLAS teams at the LHC successfully detected the much awaited Higgs boson. The obvious problem with the SM is that it is yet to include gravitational interaction, but it also encounters further difficulties such as:

- Planck scale unification - not even including gravitational interaction, it is desirable to have unification of the weak, strong and EM forces at the Planck scale, but the coupling constants of the three forces fail to unify at this scale.
- The massive neutrino - the SM expects a massless neutrino, but experiments have shown that neutrinos can oscillate between lepton flavours. This is only possible if the neutrino is massive.
- Hierarchy problem - There is no current explanation for why the weak force is 10^{32} times stronger than gravity. Furthermore, the SM cannot explain the hierarchy problem that arises when calculating the fermion loop correction to the Higgs mass.
- Dark matter and dark energy - from observations (such as rotation) of gravitational bodies, it is proposed they consist of a significant amount more matter than is observable. The SM does not propose any suitable candidate for a dark matter candidate.
- CP violation - the standard model predicts that under maximal CP violation, the maximum discrepancy between matter and antimatter could be a ratio of 10^{-10}, whereas current observations predict instead that it should be 10^{-20}.

Therefore we need to look for a theory beyond the SM that can provide answers to these fundamental questions.

1.2 Supersymmetry theory

Supersymmetry (SUSY) theory states that every SM particle has a SUSY partner with which its spin differs by one half. SUSY relates fermions and bosons, and says that each particle from one group has a partner in the other - a superpartner. Ideally we would like SUSY symmetry to be unbroken, but considering that these superpartners have not yet been seen in nature, we know that spontaneous symmetry breaking must occur if SUSY is to exist. This would allow superpartners to differ in mass from their SM counterparts. SUSY suggests an elegant solution for some of the SM issues mentioned above. For example, spontaneous symmetry breaking may be a solution to the gauge hierarchy problem. Similarly, Planck-scale quantum corrections cancel between partners and superpartners, avoiding fermionic loop corrections to the Higgs mass. Still further, SUSY unifies the weak, strong and EM forces at 10^{16} GeV. SUSY also expects massive neutrinos.

There are a number of different SUSY models with varying numbers of parameters. The Minimal Supersymmetric Standard Model (MSSM) utilises 105 parameters, whereas a theory like the Constrained Minimal Supersymmetric Standard Model (CMSSM) has only 4 (and a phase). For a comprehensive guide to SUSY, see [1].

2 Searches for SUSY

After a quark or gluon is produced in a reaction, it quickly fragments into a spray of hadrons (hadronisation) - a jet. By looking at the energy, E, transverse momentum, p_t , and other jet attributes, we can work backwards to form an idea of the particle that produced it.

The CMS detector is a general purpose detector which is comprised of three sections: the tracking section, the calorimeter and the muon tracker. In this project we will study signal and background events simulated using Monte Carlo methods (MC). At generator level we have a pure process, like the ones shown in 1, that are purely simulated. However, to make these simulations realistic, these events are processed by a simulation of the CMS detector and the full event reconstruction chain is applied, as it would be performed on data.

2.1 Interesting quantities

In order to distinguish BSM processes from SM processes, we want to look at some useful generator level quantities such as:

- Transverse momentum, p_t - this is the component of the momentum in the transverse plane to the beam line.
- Missing transverse energy, MET - in hadron colliders, the initial momentum of the colliding partons is unknown owing to the fact that the individual constituents within the hadrons share the energy between them, which we cannot measure. The energy of the particles in the direction transverse to the beam axis is originally zero. We can then check for any MET by taking the absolute value of the negative sum of all visible reconstructed particles in the transverse plane after the collision.
- Minimum pairwise mass, m_{min} - the three subjets with the highest p_t are taken pairwise and the invariant masses of the combinations are calculated. The minimum pairwise mass is the mass of the pair with the lowest invariant mass. In top-tagging, the minimum pairwise mass should reconstruct the W mass.

2.2 Jet algorithms

Jets are not always so easily defined, and there can be much pile-up to sort through. Pile-up is when the detector is affected by several events simultaneously. Jet clustering aims to reduce the complexity of the final state of the process by simplifying many hadrons into far fewer jets. Jet algorithms are a set of rules used to group together particles into jets. There are two main types of algorithm: cone algorithms and sequential-clustering algorithms.
The former is used more frequently in hadron colliders and typically involves the use of more

parameters than the latter. In a cone algorithm, the most widely used of which are iterative algorithms, we take the most energetic particle in the hadron spray - the seed (the particle with the leading momentum). Then we sum the momenta of all the particles surrounding the seed within a cone of some radius R with azimuthal angle ϕ and pseudorapidity η (or rapidity y). It is unavoidable that there arise such problems as overlapping cones, for which there are subsequent various methods to deal with the issue, such as progressive removal (remove all seeds from the event once a cone is formed) and split merging (merging a pair of cones if a significant portion of the transverse momentum is shared by particles in both cones).
Sequential-clustering algorithms, such as the k_t, anti-k_t and Cambridge-Aachen (CA) algorithms, are currently used by the CMS experiment. Sequential-clustering algorithms are based on two distance measures: the distance d_{ij} between two particles i and j, and the distance d_{iB} between any particle i and the beam B. The algorithm used to re-cluster jets in this project is the anti-k_t (AK) algorithm, where the distance measures are given by [2].

$$d_{ij} = min\left(p_{ti}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}\,, \tag{2.1}$$

$$d_{iB} = p_{ti}^{2p} \tag{2.2}$$

with

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2\,. \tag{2.3}$$

The parameter p specifies whether the reclustering is more likely to form around a hard particle or a soft particle. A hard particle is one that is more likely to come directly from the decay, and hence have a higher p_t. A soft particle is more likely to be a secondary particle in the reaction and hence have a lower p_t. In the AK algorithm, $p = -1$, which favours reclustering around hard "seeds", whereas $p = 1$ for the k_t algorithm.

3 Data analysis

3.1 Aim of this project

This project aims to investigate the use of jet substructure for a SUSY signal, to look for correlations in variables and to see whether, using Monte Carlo simulations, we see patterns we would expect. In this project we focus on gluino-gluino production of top quarks using the signal T1tttt_mGo1500_mChi100 and background $t\bar{t}$.

In order to incorporate SUSY into the SM, we need to double the number of particles. Considering the resulting large number of parameters, this means that SUSY is a difficult theory to search for, which is why we decide to search for a certain signature. In the signature we looking for, the main source of background events arises mainly from gluon gluon to $t\bar{t}$, $gg \rightarrow t\bar{t}$, which is a SM process and is shown in 1(a).

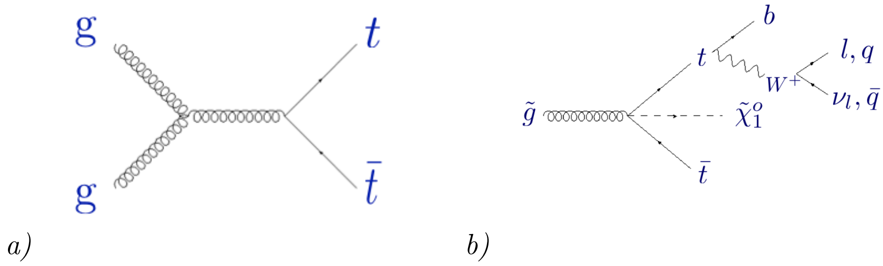

Figure 1: *a)* shows a Feynman diagram of the SM process $gg \to t\bar{t}$. *b)* shows a Feynman diagram of one branch of the SUSY process $gg \to t\bar{t} \to W, b$.

Here we would see one top, and one anti-top. On the other hand, in the simplified SUSY process we would expect to see

$$\tilde{g}\tilde{g} \to t\bar{t}\ (\to ....) + \tilde{\chi}_1^o\,,\ t\bar{t}(\to b, W \to l, \nu_l) + \tilde{\chi}_1^o\,.$$

Shown in 1(b) is one branch of this process when we examine the decay of just one of the two gluinos. This means that in our signal, we would expect to see two tops and two anti-tops (and consequently four W-bosons). It is this signature process that this project aims to inspect.

We are interested in identifying hadronically boosted top quarks from gluino-gluino production, and W bosons resulting from their decay, and tagging them. We can use multiple techniques in order to do this, like using the CMS top-tagger, W-tagger, and N-subjettiness.

3.2 Matching and Top and W tagging

To match the appropriate tops and W's to MC truth tops and W's, we can check their order in the decay chain and at the particles they were produced from. We do this by checking the particle's ID and demanding that its mother and grandmother particles also have a certain ID. A jet is matched to an MC particle if a jet in detector reconstruction could be assigned to that particle at generator level. Once we have matched jets, we can check properties of the reconstructed jet that might separate it from quark/gluon jets.

In order for jets to be tagged as top jets using the CMS top-tagging algorithm, we require that the jets have a mass close to the top mass; that there be at least 3 subjets; a p_t of at least 350 GeV and a minimum pairwise mass close to that of the W boson. In CMS W-tagging, if the original W< 0.2 away from the main axis, we can match it to the W. If it is further away than this, we match it to a quark or gluon (i.e. it is an unmatched jet). An in-depth report on the identification techniques of the CMS collaboration in the area of boosted W tagging and top-tagging can be found in [3] and [4], respectively.

The efficiency of this process is given as a ratio of those matched and tagged to those just matched. This gives us how likely we are to tag the appropriate particle.

3.3 Boosted objects

ΔR decay

The four tops resulting from gluino-gluino production can decay either leptonically or hadronically, or in a mixture of these two, but this project does not go onto cut on these variables to specify the end state of the process. The following plots show only hadronic tops.

If the top quark is very boosted, the decay products will tend in the same direction (in one jet), and hence be closer together. In comparison, if the top quark decays while it is stationary, the decay products are more likely to go off in different directions, and we would see three jets - the b jet and the $W \to q\bar{q}$ jet. If the top is slightly boosted then we are likely to see two jets.

ΔR decay refers to the distance between the decay products and the main jet axis, and we find it by calculating the four vectors of the decay products. We are interested in the maximised distance, i.e. that of the constituent the furthest from the main axis. By plotting this against p_t, we can see the correlation between how boosted the jet constituents are, and how close together they are, and from this get a more precise idea of the cone size that it might be suitable to use.

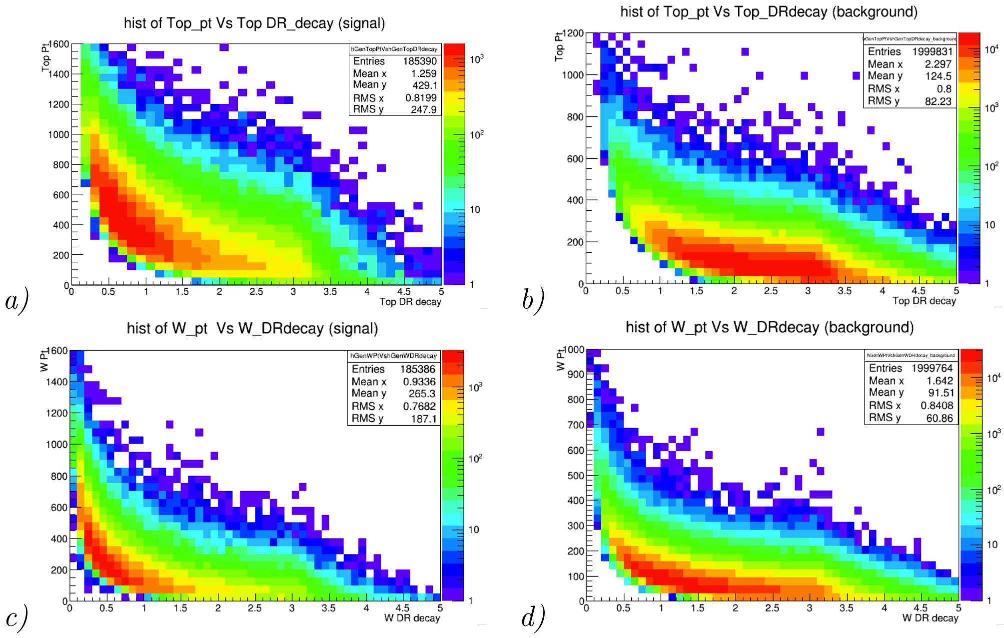

Figure 2: *a)* (and *b*) show histograms of p_t against Δ R decay for MC tops for signal (and background) . The same is shown in *c)* and *d)* but for MC W's.

From figure 2 It is clear from *a* and *c* that the higher the p_t, the closer the constituents are together, which we would expect in a boosted jet. Histograms *b* and *d* show nicely that the decay products for the background particles are further apart, and have less energy, as we would expect. Current searches involving hadronic top quarks and W bosons often utilise jets with a cone size of 0.8 or larger. However, we can see that the decay products matched to MC W's are closer together than the decay products matched to MC tops. It therefore may make more sense to separate out the signal and background for W jets even further by decreasing the cone size used.

We can then justify that, in the following analysis, we will be using a cone size of $\Delta R = 0.4$ when looking at W jets, and a cone size of $\Delta R = 0.8$ when looking at top jets. In the following paragraphs we refer to these as AK4 and AK8, respectively.This is because when identifying a top jet, it is necessary to include all the decay products so we want a larger cone size, whereas as a W-jet is more boosted, it is more meaningful to search for the decay products in a smaller cone size.

Mass and p_t studies

By looking at a plot of the jet mass against the transverse momentum for the entire AK8 signal, we can see that with increasing p_t, top and W decay products start to merge around their respective masses of 175 GeV and 80 GeV. It must be mentioned that when talking about mass, it is meaningless to look at the original mass of the jet as this depends too much on the pile-up in the event. We want to use 'grooming' algorithms to get rid of the additional energy that doesn't come from the particle, such as pruning, trimming and filtering. In other words, we want to remove the softer components in the reaction to leave behind the constituents resulting from the hard scattering. In this analysis we will use the trimmed mass, which uses the k_t algorithm to recluster subjets of a size R_{sub} from the jet.

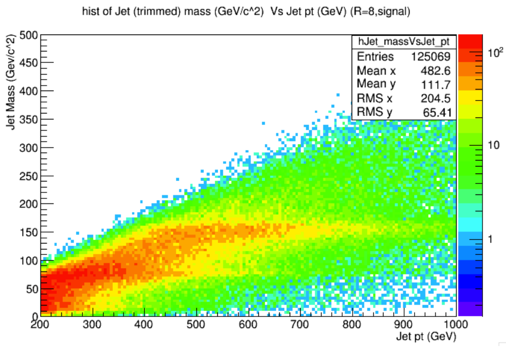

Figure 3: Histogram showing AK8 jet trimmed mass against p_t.

In figure 3, we observe in the region $200 < p_t < 500$ that the $W \to q\bar{q}$ merges first. Then at values between $500 < p_t < 800$ we see the b-jet also merging around the mass of the top quark.

After checking generator level quantities we move on to reconstructed jets and are interested in matching them to MC truth tops and W's. The first quantities we match are the (trimmed) mass and p_t, projected onto the x-axis, shown in figure 4.

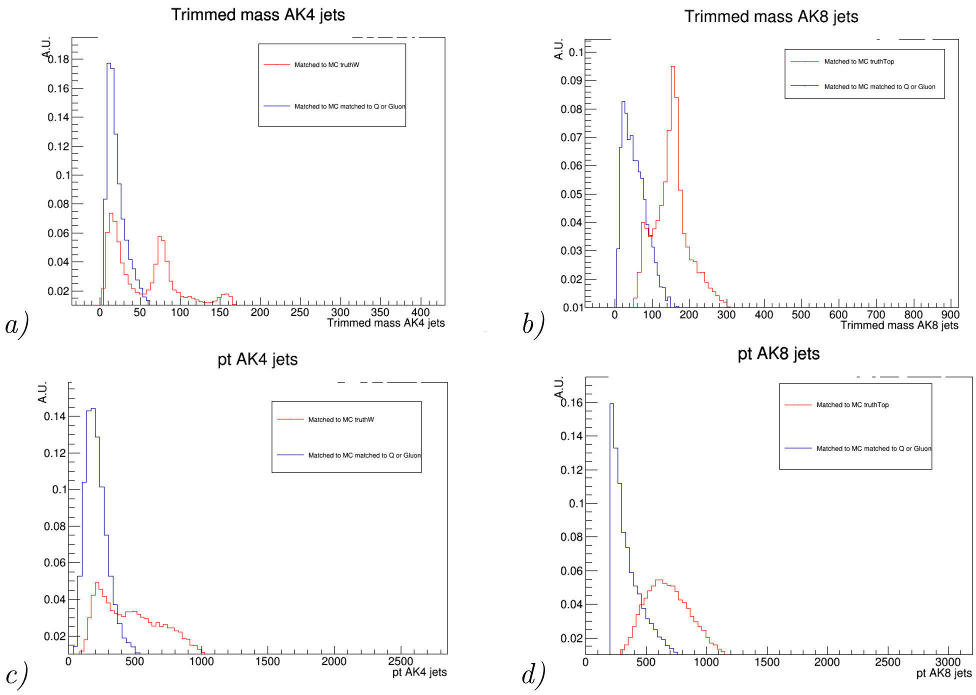

Figure 4: *a)* and *b)* show the 1D histograms of the trimmed mass of jets with $\Delta R = 0.4$, and $\Delta R = 0.8$ respectively. Below, *c)* and *d)* show 1D histograms of the p_t of jets with $\Delta R = 0.4$, and $\Delta R = 0.8$ respectively. All plots show overlays of jets matched to either a top/W alongside the jets matched to a gluon/quark, and are also normalised.

As can be seen from *a)* there is a defined peak around 80 $\frac{GeV}{c^2}$, the mass of the W boson, where the jet has been matched to an MC truth W. We can see that the jets not matched to a W have a mass peak much closer to 0 $\frac{GeV}{c^2}$, owing to the fact that quarks/gluons are closer to being massless. We can see the same pattern in *b)* where instead the jets matched to MC truth tops have a mass around 170 $\frac{GeV}{c^2}$. Again, in *c)* and *d)* we see that the p_t of the jet not matched to a W or top is significantly shifted towards the left, showing us the remaining p_t that is of real interest when seeking our signature process.

We can similarly show these correlations by plotting the mass and p_t against each other like in figure 5.

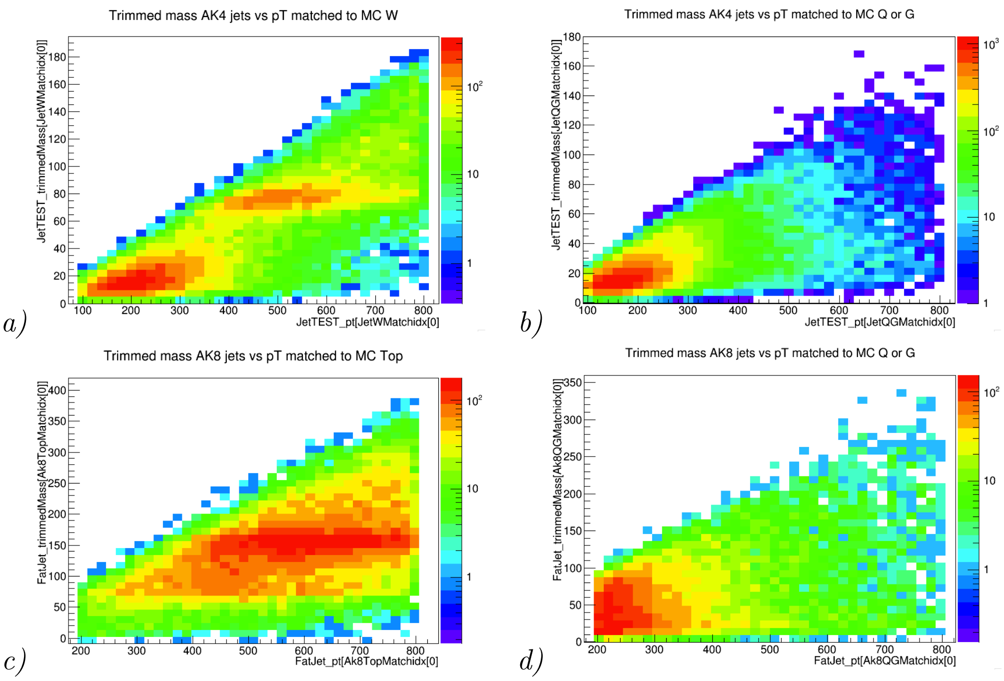

Figure 5: *a)* and *b)* show 2D histograms of trimmed mass Vs p_t for AK4 jets matched and unmatched to an MC truth W jet (respectively) . The same is shown in *c)* and *d)*, instead matching AK8 jets to MC truth tops. In all plots there is a requirement $100\frac{GeV}{c} < p_t < 800\frac{GeV}{c}$. Further, JetMatchidx[0] indicates that only the leading order matched jet is shown in the plot. Note that fatjets are heavy jets with larger particles.

In *a)* we can see a higher density of events with a higher p_t around the mass of the W boson. In *c)* there is a high density of events around the mass of the top quark, but with larger spread of p_t. This pattern indicates that the tops decaying from the gluinos are boosted by different amounts. In both *b)* and *d)* we see a cluster of low p_t, low mass events that are neither matched to tops nor W's.

N-subjettiness

N-subjettiness describes a jet shape, and is useful in boosted object determination. It is an effective measure for identifying how the jet energy is divided up into subjets, and is calculated by taking the ratio of $N+1$ and N, which gives us how likely a jet is to have $N+1$ subjets. For these studies, jets are again spatially matched to generator level tops and Ws. A detailed look into the use of N-subjettiness is given in [5]. N-subjettiness is defined in terms of the variable τ_N, which is found by summing over the constituents within a jet and minimising the distance to the nearest subjet axis. As the name implies, N defines the number of subjets. τ_N takes the value $0 \leq \tau_N \leq 1$, with a small τ_N indicating constituents are close to the subjet axes, and a large τ_N indicating the constituents are spread out.The jet shapes best describing tops and W's are given by

$$\text{W-subjettiness} = \frac{\tau_2}{\tau_1} \ , \tag{3.1}$$

$$\text{Top-subjettiness} = \frac{\tau_3}{\tau_2} \ . \tag{3.2}$$

It happens that (3.1) well describes a 2-pronged objet such as a W-jet, whereas (3.2) better describes a three-pronged objet such as a top-jet.

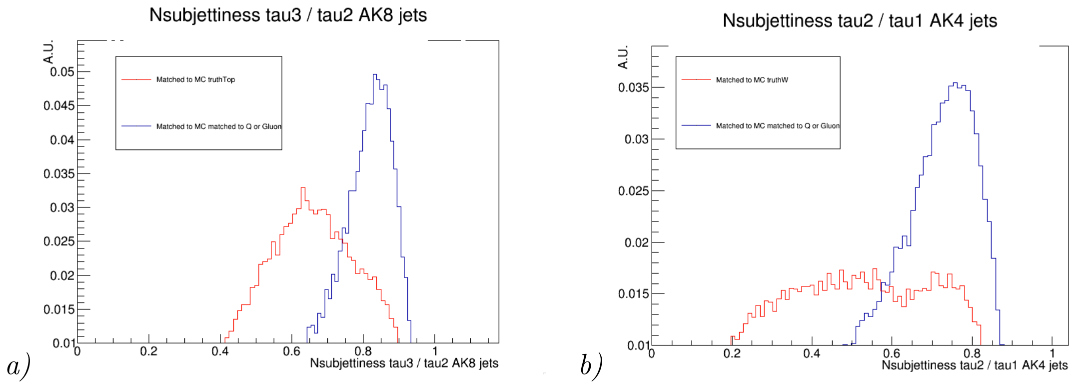

Figure 6: *a)* and *b)* show the 1D histograms of the x-projection of the top and W subjettiness respectively.

In figure 6, the curves of the jets matched to MC truth tops (W's) are smaller which indicate that they are more likely to contain 3 (2) subjets than those matched to quarks/gluons. The curves of both MC particles matched to quarks/gluons are shifted to the right, towards a value of 1. From the red and blue curves we can determine that a cut on on the subjettiness (0.6 for tops and 0.5 for W's) would help to separate signal and background. While increasing the cut value increases the identification efficiency, it also yields a higher false positive rate, i.e. one gets more background classified as signal.

4 Conclusions and Outlook

In this project, an investigation into some of the variables describing jets has been carried out. Histograms of trimmed mass and transverse momentum have shown what we would expect, and have given a taste into the use of matching and how matching enables us to differentiate between the initial noise of the collision, and the SUSY signature we are searching for. As an extension, it could be interesting to compare some of the plots here with new ones done using different mass-grooming techniques and to implement cuts to see how these change results. The next step in this project would be to inspect the efficiencies between the rate of jets matched and tagged, and those matched. We have seen that, although a cone size of 0.8 is used by default in both top and W tagging, it could yield more information were we to reduce the cone size when looking at W tagging, and an investigation into this is suggested.

Acknowledgements

I would like to thank the DESY-CMS group for welcoming me to study this summer, and especially to Claudia Seitz and Isabell Melzer-Pellmann in the SUSY group for taking me under their wings, and imparting their knowledge.

References

[1] S. P. Martin, "A Supersymmetry primer", arXiv:hep-ph/9709356, [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[2] G. P. Salam, "Towards Jetography", *Eur. Phys. J.* **C67** (2010) 637–686, arXiv:0906.1833.

[3] **CMS** Collaboration, V. Khachatryan *et al.*, "Identification techniques for highly boosted W bosons that decay into hadrons", *JHEP* **12** (2014) 017, arXiv:1410.4227.

[4] **CMS** Collaboration, C. Collaboration, "Boosted Top Jet Tagging at CMS", 2014.

[5] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **03** (2011) 015, arXiv:1011.2268.