Kinematic Analysis of t-tbar events in CMS

Chris Parkinson

Supervisor: Prof. Dr. J. Mnich

September 16, 2008

Abstract

The top quark is the least measured, and yet the most interesting, quark of the standard model. In order to measure its properties, accurate and reliable reconstruction of semi-leptonic t-tbar events must be achieved. This investigation analysed both angular separation and event shape at generator level in order to find possible inputs for a likelihood function, in order to reduce the effects of the combinatorial background. It was found that the ΔR value for combinations of W-jets and b-jets was suitably different for correct and incorrect combinations to allow use in a likelihood function. In addition, separating the event into the two decay chains, via an intersecting plane, also gave suitable results for use in a likelihood function.

1 Introduction

One hundred metres underground, beneath the city of Geneva, lays an accelerator approximately 27km long called the Large Hadron Collider (the LHC). This accelerator is a huge synchrotron system, designed to boost two beams of protons to the staggering energy of 7TeV, and then allows the beams to collide at four points within the ring. This system is the largest and most complex machine ever built by man, with the primary aim of searching for new particles at a previously unexplored energy scale.

At each of the four collision points, proton-proton interactions will occur, allowing us to observe unseen physics at higher energies and higher luminosities than ever before. A huge range of physics will be available for study in these interactions, such as the famous Higgs searches, however other important physics will also be studied. One such topic is that of top physics, in particular, the study of top anti-top pair production.

However, the study of these interactions is not trivial. Vast and intricate detector systems are required to sense, and measure the properties of, the particles created in the interactions. One such detector system is the Compact Muon Solenoid (CMS) system, an 'all-purpose' detector located at one of the four crossing points within the LHC. The detectors combine leading-edge technology with novel computing and analysis techniques to allow the conversion from mysterious particle interactions into observable and understood physics processes. A sketch of the CMS detector, showing the various subdetectors, can be seen in figure 1

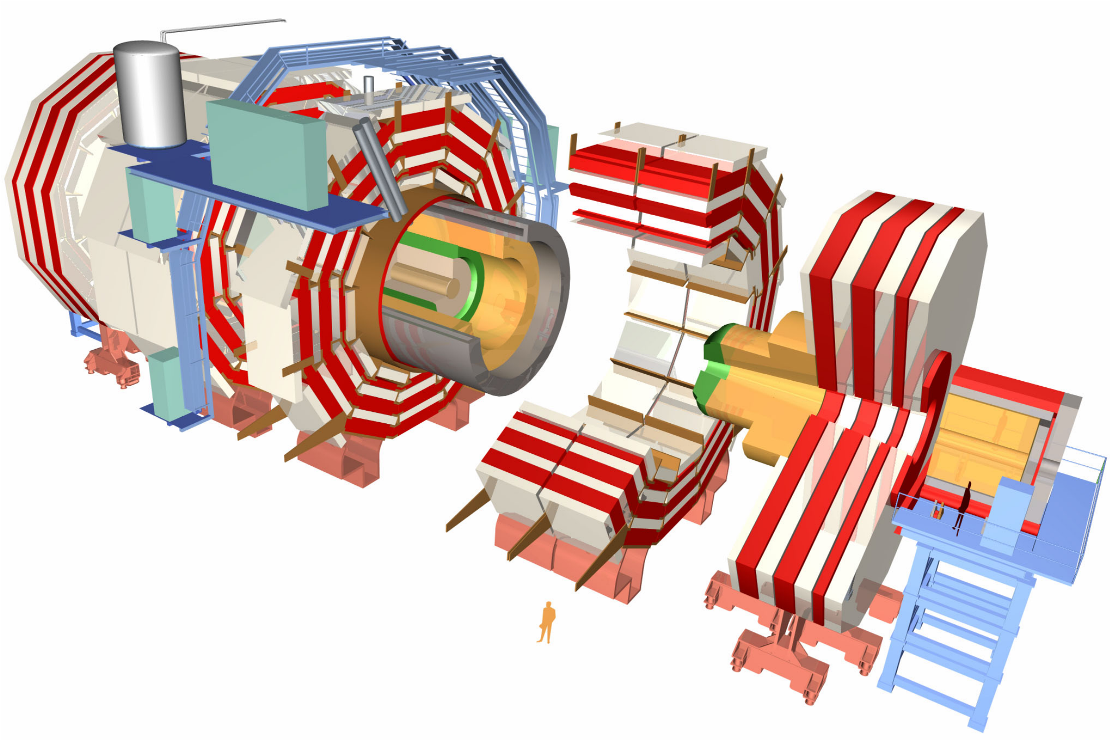

Figure 1: Sketch of CMS displaying the various detector subsystems

1.1 Top Physics

One type of interesting event that will be detected in CMS are top physics events, which are classified by the production of a top quark in the initial state. This top quark rapidly decays, creating decay products that either decay further, or hadronise, creating events that detectors such as CMS can then attempt to detect and measure.

One particularly interesting type of top event is the production of a t-tbar pair. At the LHC accelerator, this event will be dominantly produced via gluon fusion, the interaction of two gluons. This event will evolve with the almost immediate decay of the two top quarks into a lighter quark (usually a b-quark due to the large CKM matrix element) and a corresponding W particle. The W particles will further decay in one of two ways, hadronic or leptonic, producing either a quark anti-quark pair or a lepton lepton-neutrino pair respectively. In addition, any quarks that are produced during the decay will quickly hadronise, forming a jet of hadrons, as quarks cannot exist as free particles.

This then leaves three dominant final state systems from a typical t-tbar event. The first is 'Fully Hadronic', in which case both the W particles decay via a hadronic mode, resulting in six jets in the final state. This event type is relatively difficult to reconstruct due to the large number of jets present in the final state. Additionally, the event may evolve into a 'Di-Leptonic' final state, in which case both the W particles decay via a leptonic mode. This event type is also difficult to reconstruct, due to the production of two neutrinos in the final state, which cannot be seen in the detectors at the LHC. The combinatorics of combining two unseen particles correctly poses a difficult challenge. Interestingly however the third possible final state, 'Semi-Leptonic' — where one W particle decays hadronically and the other leptonically — offers a unique insight to the t-tbar event. This is due to the lower number of jets produced (four, compared to the six in a fully hadronic event) and the production of a single neutrino (compared to the two produced in a Di-Leptonic event). Thus, the semi-leptonic event is the easiest of the event types to analyse. A typical semi-leptonic event is shown in figure 2.

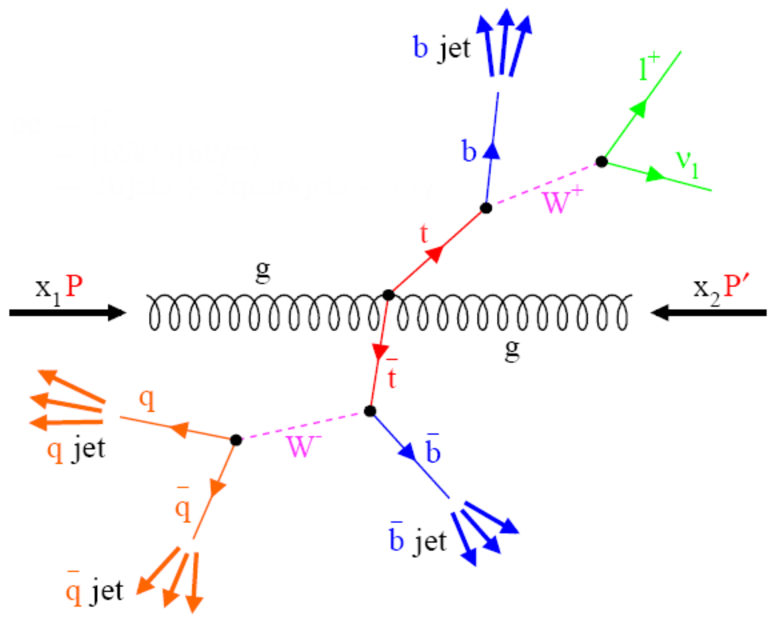

Figure 2: Simple diagram showing a typical semi-leptonic t-tbar event

1.2 Event Reconstruction

Despite the relative ease of reconstructing a semi-leptonic t-tbar event, there are still major challenges in doing so. For example, there is considerable issue with combinatorics when reconstructing an event with six final state particles, especially as four of the particles produce hadronic jets, and one further particle goes completely undetected.

In order to overcome these issues, the detectors at the LHC are designed to gain enough event information to allow varied post-event analysis. In order to identify which particles were present in the initial state, particle ID algorithms are used to identify particles. In addition, jet algorithms are used to consider whether the particles which were detected in fact have originated from a single quark. Importantly, these jet algorithms can identify whether the jets originated from a b or c-quark due to the existance of displaced vertices near the interaction point, due to the production of a relatively long-lived B meson, and characteristic energy deposits in the calorimeters. In addition, transverse momentum p_T and pseudo-rapidity η cuts are made to remove particles that are unlikely to belong to the t-tbar event, as well as analysis of missing transverse energy and momentum (E_T and p_T) to identify any undetected particles such as neutrinos.

Once the particles in the event have been identified, it is possible to analyse what type of event produced the particles. Jet number cuts may be made to exclude events that do not appear to be of the correct form, for example, four jets are created in a t-tbar event, so any event with less or more jets can be removed as being unlikely to be a t-tbar event. In addition, two of the jets are extremely likely to originate from b-quarks, as well as two of the jets being from the same W particle. This then allows a cut using the type of jets that are seen (in the t-tbar case, two b-jets and two W-jets may be used to select events).

In addition to these cuts, event shape and angular separation of particles, as well as a myriad of other variables, may be analysed to further distinguish the various events.

2 Analysis Techniques and Results

In this investigation, the event shape and angular separation of particles were analysed using a generated set of t-tbar events. The aim was to identify quantities that could be used as input to a likelihood function, to attempt to reduce the impact of the combinatoric background. Ultimately, this would aid event reconstruction of t-tbar events in CMS.

2.1 Analysis One: Angular Separation using ΔR

2.1.1 Technique - At generated quark level

The first investigation that was performed was analysis of the variable ΔR for various particle combinations of the t-tbar events.

The variable ΔR is defined by:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

where η is pseudo-rapidity, and is used to remove the effect of boosting due to the unsymmetric nature of LHC interactions. The purpose of ΔR is that it gives a measure of angular separation that is independent of the boost of the system.

The variable ΔR was used in order to differentiate between correct and incorrect combinations of the four final state quarks into the decay chains. This distinction was defined such that a combination of two quarks from the same initial top particle would be a correct combination, whereas a combination of two quarks, each of which originated from different top particles, would be considered a wrong combination. A correct combination can be seen in figure 3 (left), where all highlighted particles originate from the same top quark. An incorrect combination can be seen in figure 3 (right), where particles from both the top and anti-top quarks are highlighted.

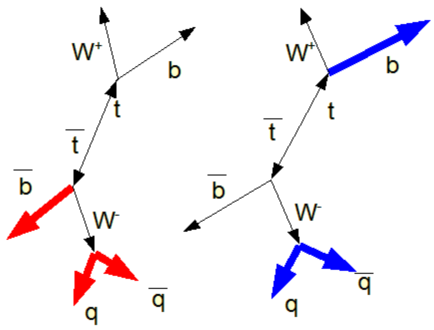

Figure 3: Simple diagram showing a correct, and an incorrect, quark combination

The particles that were analysed using this method were: W^+ particle with b-quark (correct combination), W^+ particle with bbar-quark (incorrect combination); W^+ decay quarks with b-quark (correct combination) and W^+ decay quarks with bbar-quark (incorrect combination); W decay quark one with W decay quark two — for both decaying W^+ and decaying W^-; . The method was repeated for three transverse momentum (p_T) cuts on the relevant final state quarks. Initially, the results for no cut were analysed. For comparison, a soft cut of 25 GeV, and a hard cut of 40 GeV, were also made.

2.1.2 Results

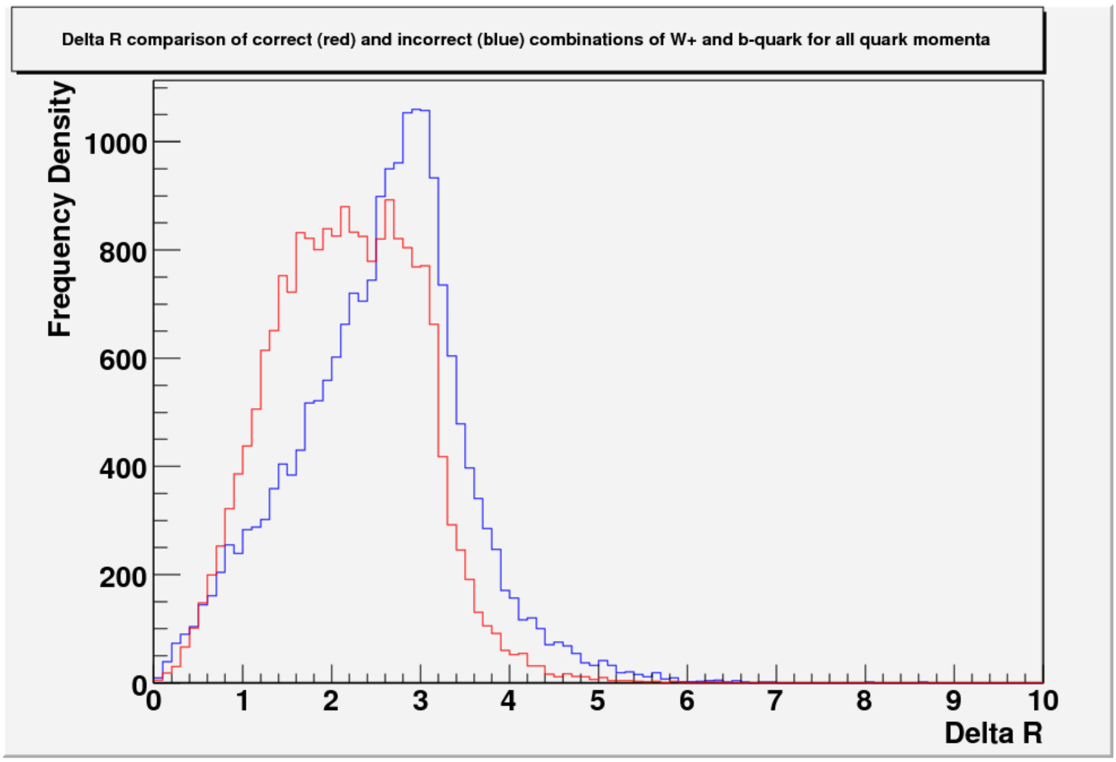

Figure 4: Delta R separation of correct (red) and incorrect (blue) combinations of W^+ and b-quark, for all particle p_T

Figure 4 displays the ΔR distributions for correct and incorrect combinations of W^+ and b-quarks with no transverse momentum cuts. The correct combination can be seen as the relatively wide distribution centred around a ΔR value of approximately two (red). The incorrect combination can be seen as the relatively thin distribution centred around a ΔR value of approximately three, relating to the roughly back-to-back nature of the incorrect combinations. Of particular interest is the relatively larger values of ΔR for the incorrect combinations, particularly at the peak at three, and the long tail of values from three to six. Between values of 1 and 2.5, there is a relative abundance of correct combinations. The differences in the two distributions are what can be used as input for a likelihood function.

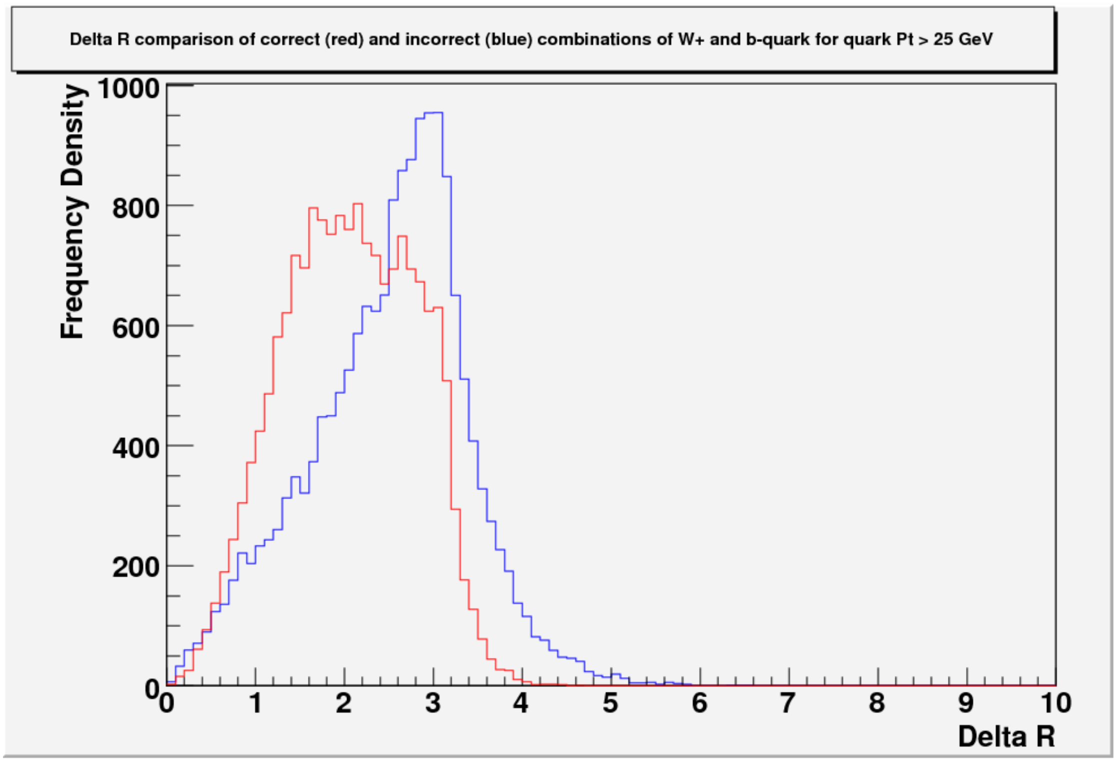

Figure 5: Delta R separation of correct (red) and incorrect (blue) combinations of W^+ and b-quark, for particle p_T greater than 25 GeV

Figure 5 displays the ΔR distributions for correct and incorrect combinations of W^+ and b-quarks with a particle transverse momentum cut of 25 GeV. The same basic distribution can be observed in this plot as seen for that with no momentum cut. A reduction of correct combinations in the ΔR range of $2.5 \rightarrow 6$, however, gives a clearer separation of correct from incorrect combinations. Interestingly, the number of events is relatively unaffected by the soft momentum cut.

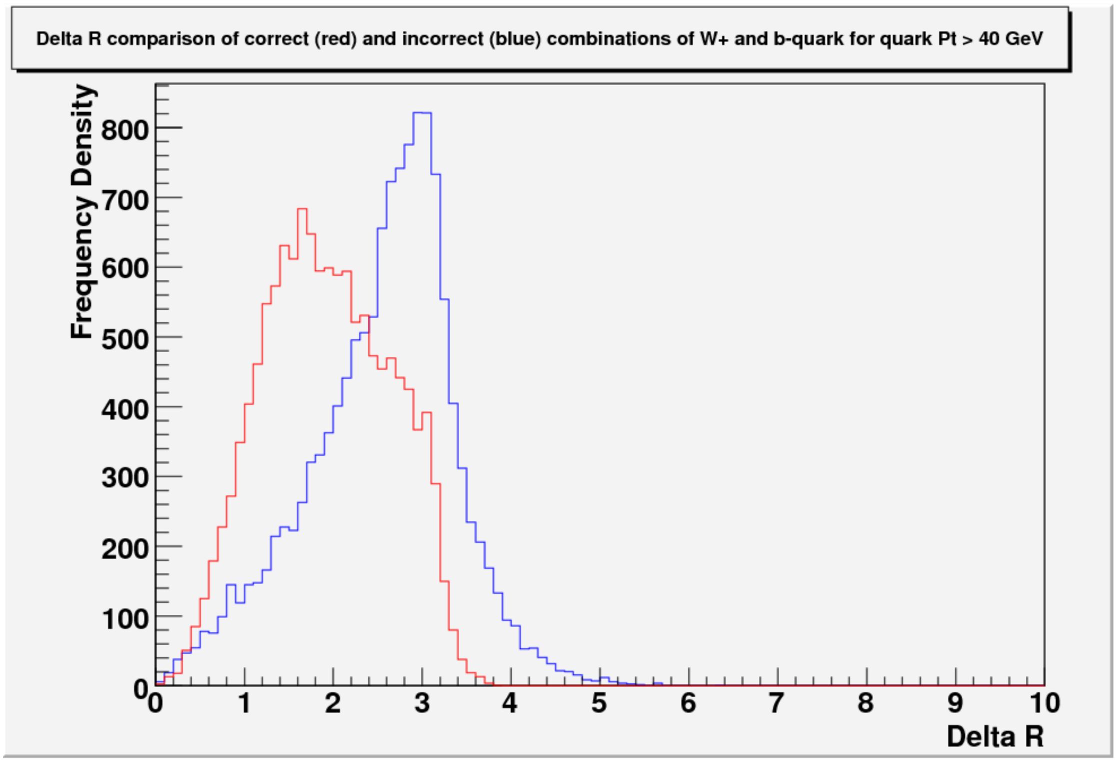

Figure 6: Delta R separation of correct (red) and incorrect (blue) combinations of W^+ and b-quark, for particle p_T greater than 40 GeV

Figure 6 displays the ΔR distributions for correct and incorrect combinations of W^+ and b-quarks with a particle transverse momentum cut of 40 GeV. The same basic distribution can be observed in this plot as seen for that with the soft momentum cut. Once again, the number of correct combinations with ΔR values greater than 2.5, and a slight sharpening of the blue (incorrect combination) peak, gives better separation between the two distributions. Unfortunately, the hard momentum cut of 40 GeV has reduced the number of events in the distributions.

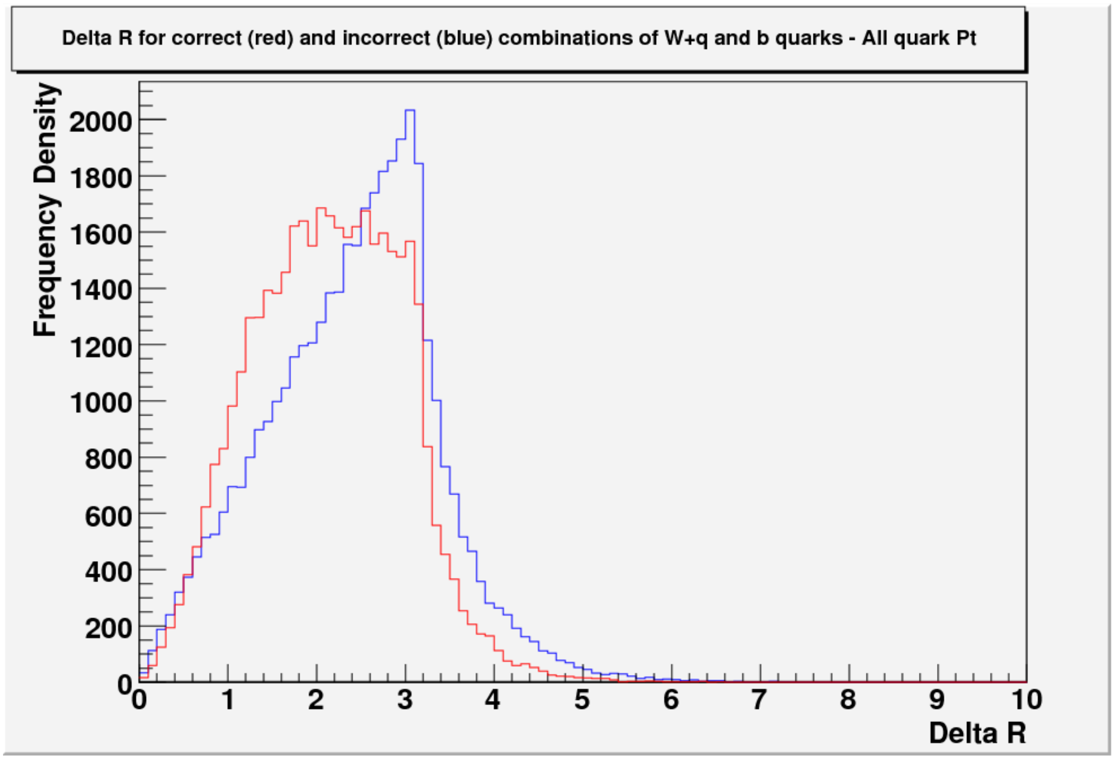

Figure 7: Delta R separation of correct (red) and incorrect (blue) combinations of W decay and bottom quarks, for all quark p_T

Figure 7 displays the ΔR distributions for correct and incorrect combinations of W decay quarks and b-quarks, with no cut on particle transverse momentum. The correct combination distribution can be seen in red as the wide peak at ΔR values of approximately two. The incorrect combination distribution can be seen in blue as the relatively sharp peak at a ΔR value of around three. In this histogram, there is again an abundance of incorrect combinations in the ΔR range $2.5 \rightarrow 6$, and an abundance of correct combinations in the ΔR range $0.8 \rightarrow 2.5$.

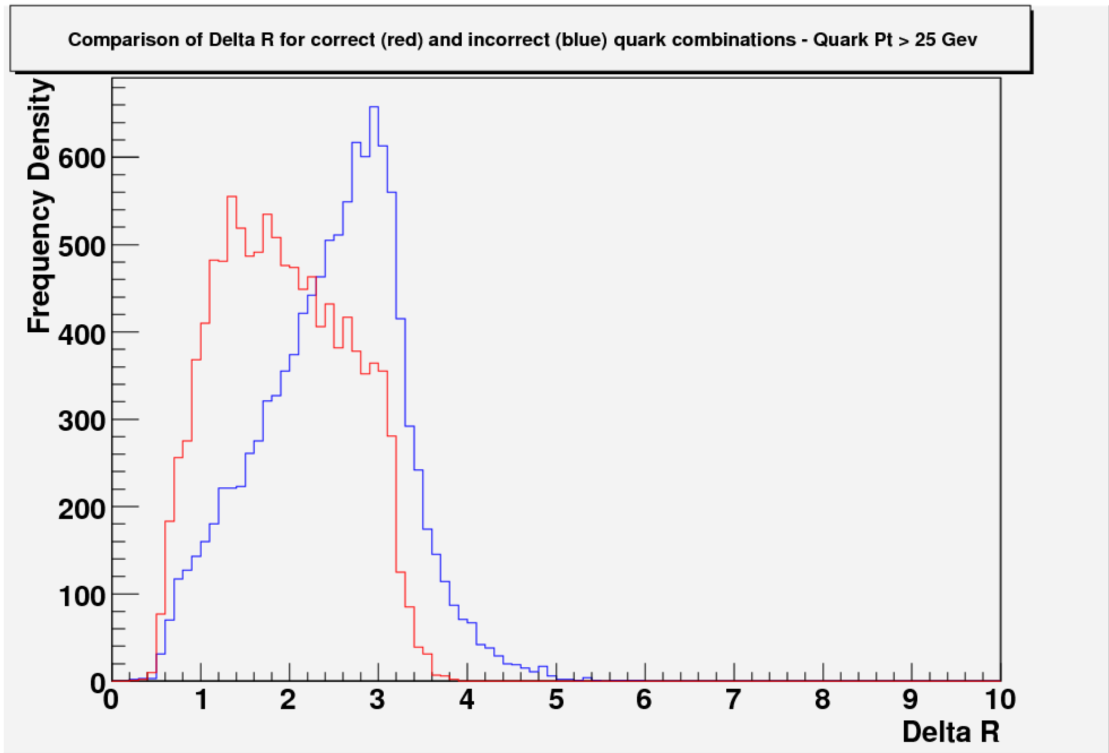

Figure 8: Delta R separation of correct (red) and incorrect (blue) combinations of quarks, for quark p_T greater than 25 GeV

Figure 8 displays the ΔR distributions for correct and incorrect combinations of W decay quarks and b-quarks with particle p_T greater than 25 GeV. The same basic distribution can be seen in this plot as the previous, with the correct combination peak at approximately one, and the incorrect combination peak at approximately three. Similarly to the plots with W^+ and b (or bbar) particles, the soft momentum cut has reduced the correct combinations at larger values of ΔR. The momentum cut has had a more noticeable effect on the number of events in the distributions, however.

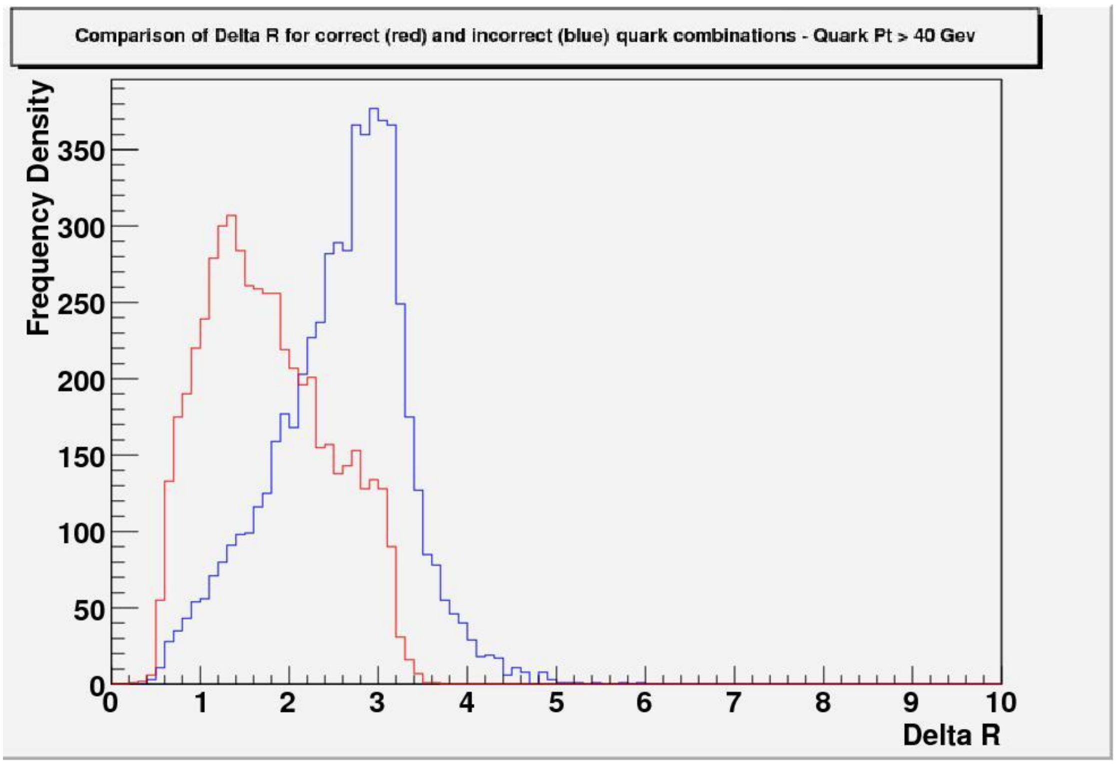

Figure 9: Delta R separation of correct (red) and incorrect (blue) combinations of quarks, for quark p_T greater than 40 GeV

Figure 9 displays the ΔR distributions for correct and incorrect combinations of W decay quarks and b-quarks with particle p_T greater than 40 GeV. Once again, the same basic distribution can be seen in this plot. The hard momentum cut has reduced the number of correct combinations dramatically for ΔR between two and three, as well as a reduction of incorrect combinations in the same region. This again leads to a greater ability to distinguish between the two distributions.

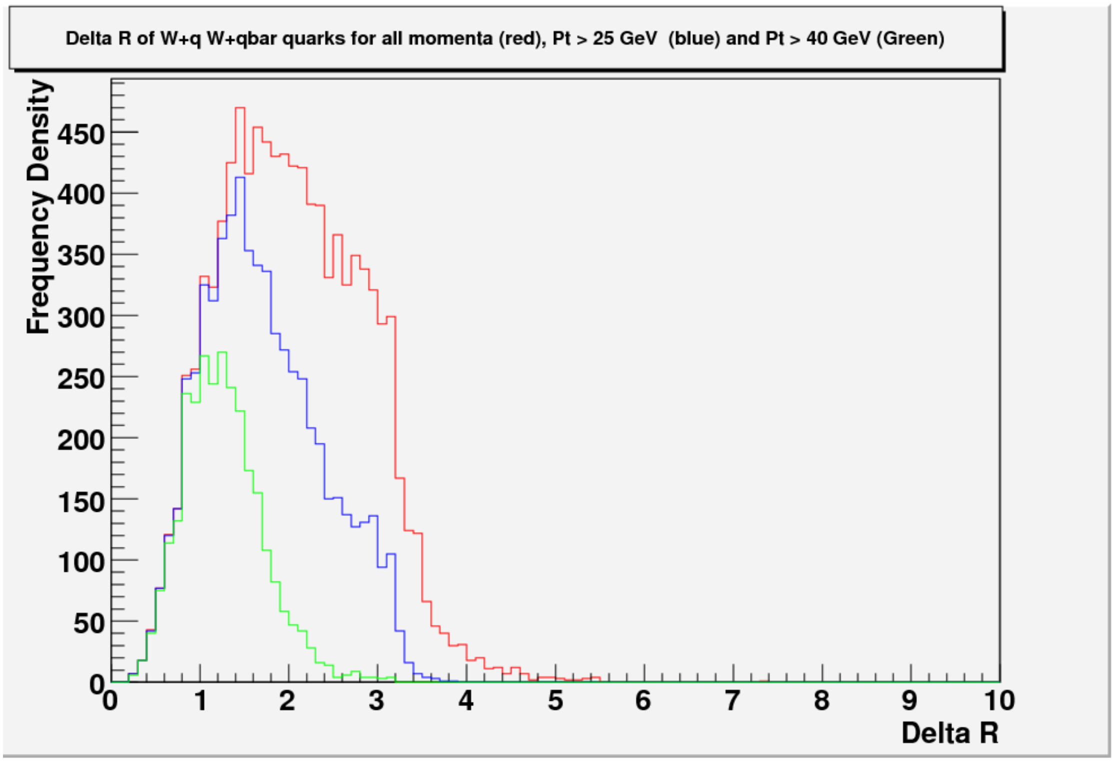

Figure 10: Delta R separation of the quark anti-quark pair from the hadronic decay of the W^+ particle, for no p_T cut (red), for quark p_T greater than 25 GeV (blue), for p_T greater than 40 GeV (green)

Figure 10 displays the ΔR distributions for the quark anti-quark pair from the hadronic decay of W^+ particles. The same distribution was plotted for three momentum cuts; with no transverse momentum cut (red), with a transverse momentum cut of 25 GeV (blue) and with a transverse momentum cut of 40 GeV (green). The red distribution can be seen to peak around the ΔR value of 1.6, and is relatively wide, with a tail leading to higher values. The blue distribution is considerably sharper, with a much more defined peak at 1.4, and again with a tail leading to higher values. The green distribution is a well defined peak at 1.2, again with a tail to higher values. The effect of momentum cutting can clearly be seen in the number of events in each consecutive plot — there are considerably less events in the green distribution than in the red — which indicates a loss of statistics for the hard momentum cut, although the sharper peak does allow more reliable analysis.

The distributions for ΔR of the quark anti-quark pairs from hadronic decays of W^- particles are identical.

2.1.3 Technique - At generated jet level

The previous technique was then extended by investigating the same quantities using generated jets instead of generated quarks. This gives a more realistic view of the event, as it is not quarks which are seen in detector systems, but the hadronic jets that they produce.

This extension has implications for the results to be obtained. Firstly, the use of jet algorithms with generated jet events immediately reduces the quality of the quark momentum reconstruction — effects such as particles which should be included in the jet being omitted, or particles that do not belong to the event being included, cause this loss of quality. This effect is shown in the simple diagram labeled figure 11, in which correctly assigned particles are shown as black lines, incorrectly assigned particles are shown as red lines, and the blue circle describes what the jet algorithm considers to be the jet. In addition, the jet algorithm being used (cone05) will remove any two jets with a ΔR value of less than 0.5 as these will be considered to be the same jet. Furthermore, there is a standard cut on jet transverse momentum (p_T) of 25 GeV on the data set to remove events without a clear t-tbar signal. This removes the possibility of creating histograms which include all jet momenta, the lowest cut made was 25 GeV. In addition, the events were required to contain four jets, two of which have a large b-tag value and two of which appear to originate from a single W particle, such that only events likely to be true t-tbar events are analysed.

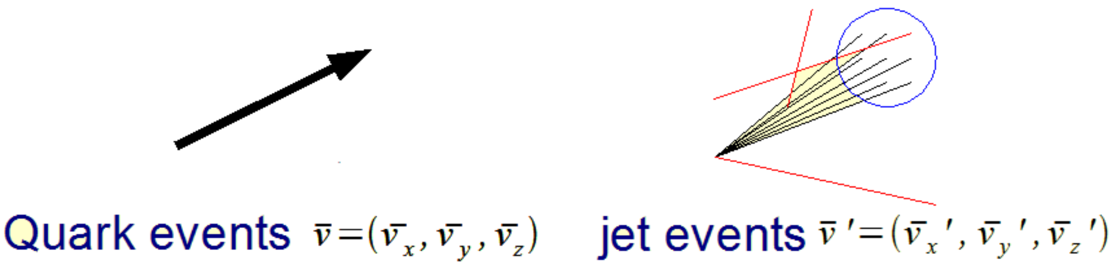

Figure 11: Simple diagram indicating the issues of dealing with jets, such as badly assigned particles, here shown in red. The jet algorithm selection is shown in blue.

2.1.4 Results

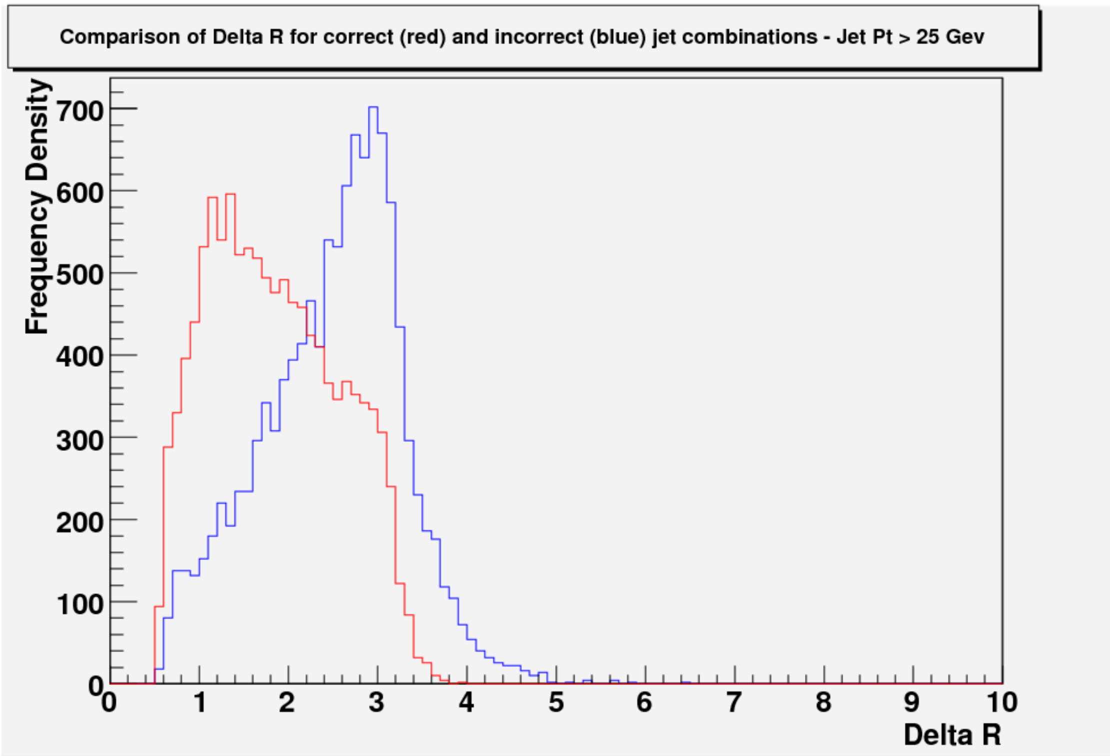

Figure 12: Delta R separation of correct (red) and incorrect (blue) combinations of W and b jets, for jet p_T greater than 25 GeV

Figure 12 shows the ΔR separation of W and b jets for correct (red) and incorrect (blue) combinations, with a p_T cut made on the jets of 25 GeV. The correct distribution is a relatively wide distribution which peaks at a ΔR value of approximately one. The incorrect distribution is a relatively sharp distribution that peaks at a ΔR value of three, reflecting the back-to-back nature of the incorrect combination. Both distributions have a tail leading to higher values of ΔR. Between values of 0.5 and 2.2, there are considerably more correct combinations than incorrect ones, whereas from 2.2 to 5, there are more incorrect combinations. This difference in the distributions is the quality that allows the information to be used as input for a likelihood function. Importantly, the histogram shows that the separation of the two distributions is not lost by moving to the jet level.

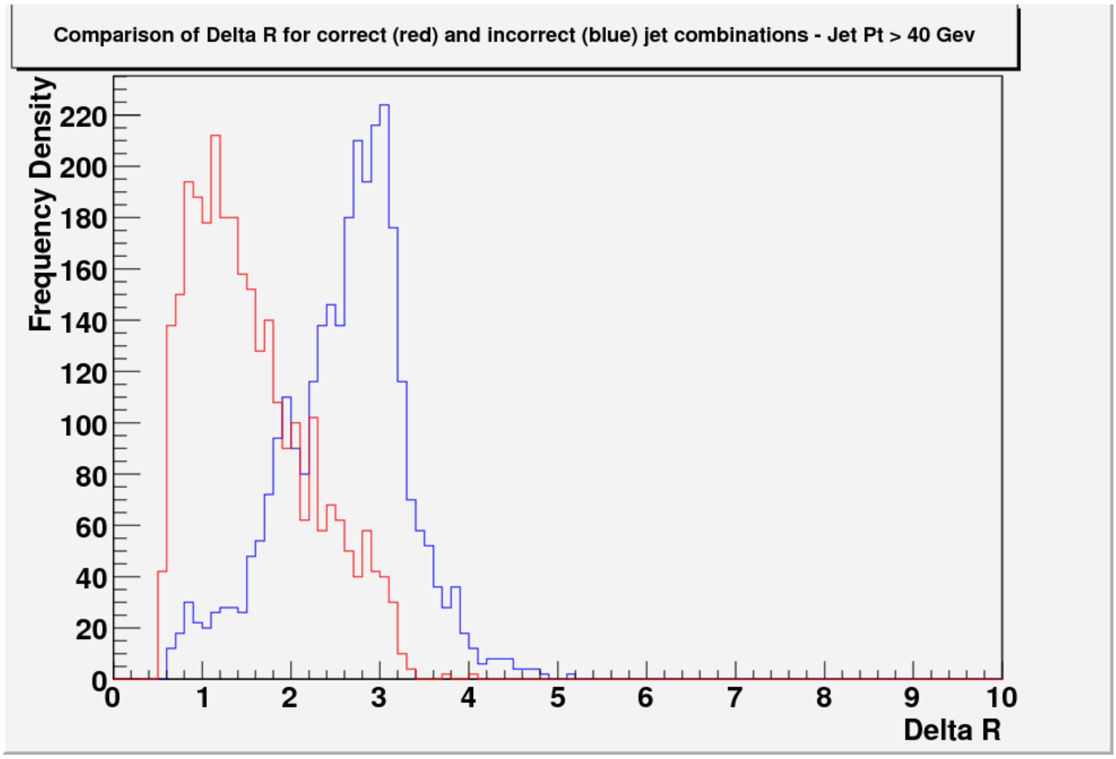

Figure 13: Delta R separation of correct (red) and incorrect (blue) combinations of W and b jets, for jet p_T greater than 40 GeV

Figure 13 shows the ΔR separation of W and b jets for correct (red) and incorrect (blue) combinations, with a p_T cut made on the jets of 40 GeV. The distributions seen in this histogram are similar to those seen in figure 12, however, the hard momentum cut has modified the distributions. The peaks of the two distributions are much more defined with the harder p_T cut, with a large reduction of correct combinations with a ΔR value greater than about 1.5, and a large reduction of incorrect combinations with ΔR lower than around 2. Importantly, the total number of events has been dramatically reduced by the harder cut, implying a loss of statistics. However, this hard cut has made the distributions quite distinct, allowing considerable separation of the two.

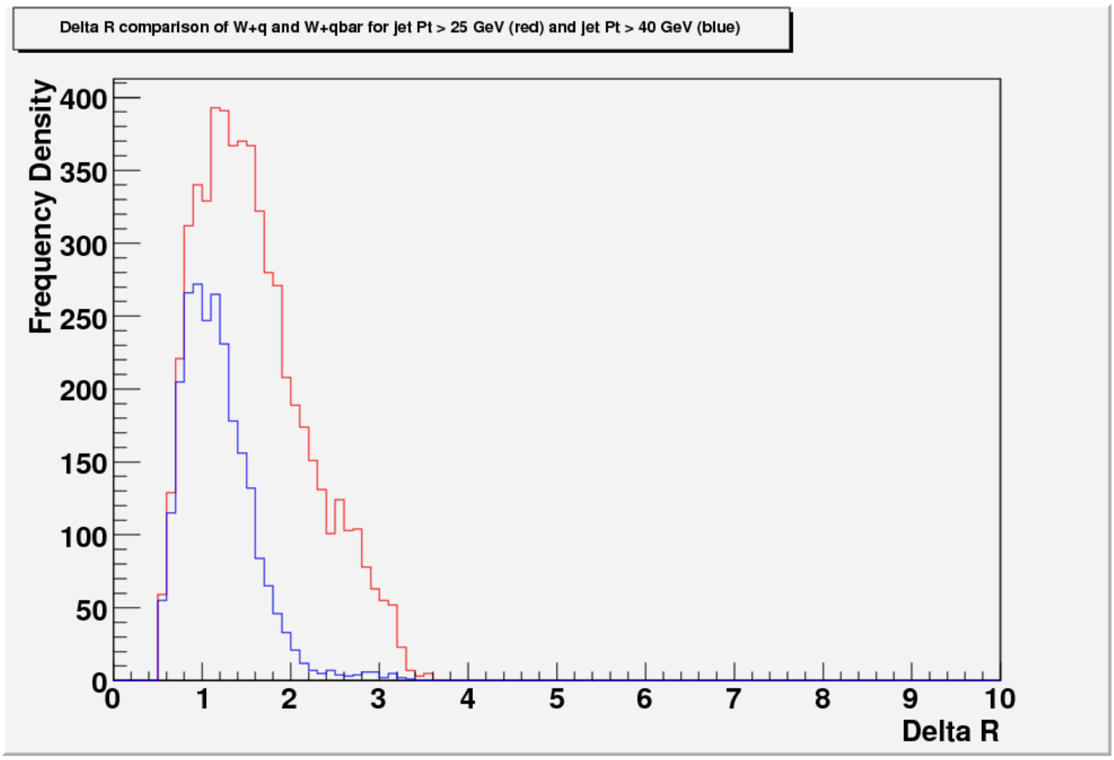

Figure 14: Delta R of quark anti-quark pair from the decay of a W^+ particle, for jet p_T cuts of 25 GeV (red) and 40 GeV (blue)

Figure 14 shows the ΔR distribution for the quark anti-quark pair produced in the hadronic decay of a W^+ particle, for jet p_T greater than 25 GeV (red) and 40 GeV (blue). The red data can be seen to be a peaked distribution, with the peak around a ΔR value of 1.3. The blue distribution is also a peaked distribution, which is relatively thinner than the red distribution, and is peaked around a ΔR value of 1. The loss of events, and thus statistics, is noticeable from the relatively small blue distribution as compared to the red distribution. Interestingly, the red distribution does not display a long tail reaching higher ΔR values, whereas the blue distribution does.

2.2 Analysis Two: Event Shape using half-space separation

2.2.1 Technique

To look at further quantities to possibly use as input for a likelihood function, the shape of the t-tbar events were investigated by separating the events in two halves.

The 'half-space' plane may be defined by a normal vector and a point in the three dimensional co-ordinate system. For the purpose of this investigation, a cartesian co-ordinate system was used, where the interaction point of the detector is defined as the origin. In order to separate the event into two halves, the plane must be defined in order to separate the decays of the initial top and anti-top quarks at the interaction point. This is not completely trivial, however, as the two top particles can essentially travel in any direction from the interaction point, which is not necessarily in the centre of the co-ordinate system.

In order to define the plane, the momentum vectors of the two top quarks were used, and the plane formed by the two vectors (the 'vector plane') was considered. The normal vector of each top momentum vector was taken, and these normal vectors were added vectorially to obtain the 'resultant' vector, which occurs in the direction directly between the top vectors and in the same plane. Thus, in order to define a plane which separates the two top decays, the normal must be defined perpendicular to the resultant vector and in the same plane (the vector plane). In order to do this, the vector (cross) product of the two top momentum vectors was taken, to essentially find the normal to the vector plane (the 'vector plane normal'). The vector (cross) product of the vector plane normal and the resultant vector was then found, to obtain the normal to the half-space plane (the 'half-space plane normal'). The origin of the co-ordinate system was used to define the plane. A diagrammatic view of the situation can be seen in figure 15.

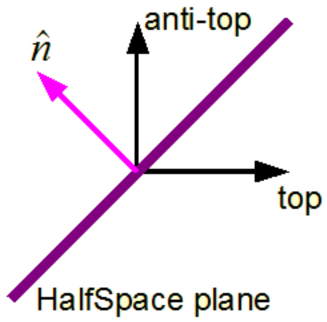

Figure 15: Simple diagram of the half-space plane alignment. This is a top-down view of the vector plane, with the half-space plane shown in dark purple intersecting the top and anti-top momentum vectors. The half-space plane normal can also be seen.

With the plane defined, it was then possible to look at which side of the half-space any particular particle was on. To do this, the scalar (dot) product of the particle's momentum vector and the half-space plane normal was calculated. It is then found that a positive value of the scalar product would indicate the particle lay 'outside' (on the same side as the normal vector, as the angle between them will be less than 90 degrees) of the half-space, whereas negative values of the scalar product would indicate the particle lays 'inside' the half-space. This is shown diagrammatically in figure 16

This process was repeated for the four final state quarks of each t-tbar event. If a W^+ decay quark, for example, was found on the same side of the half-space as the top quark, then the decay quark was considered to be in the correct half-space. If the decay

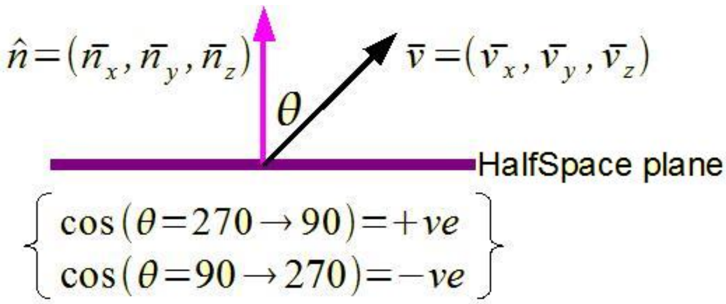

Figure 16: Simple diagram indicating how the containment of a particle can be found from finding the scalar (dot) product of the half-space plane normal ($\hat{n}$) and the arbitrary particles momentum vector ($\bar{v}$).

quark was found to be on the opposite side of the half-space as the top quark, then it was considered to be in the incorrect half-space. In order to quantify this, a value of one was assigned to any quark in the correct half-space, and a value of zero was assigned to any quark in the incorrect half-space. Thusly, the 'containment quality' can be defined as an integer quantity ranging from zero to four indicating the amount of quarks in the event contained to their correct half-space. Naturally, having all quarks in the correct half-space is admirable, so a containment quality value of four indicates a 'perfect' event.

2.2.2 Results

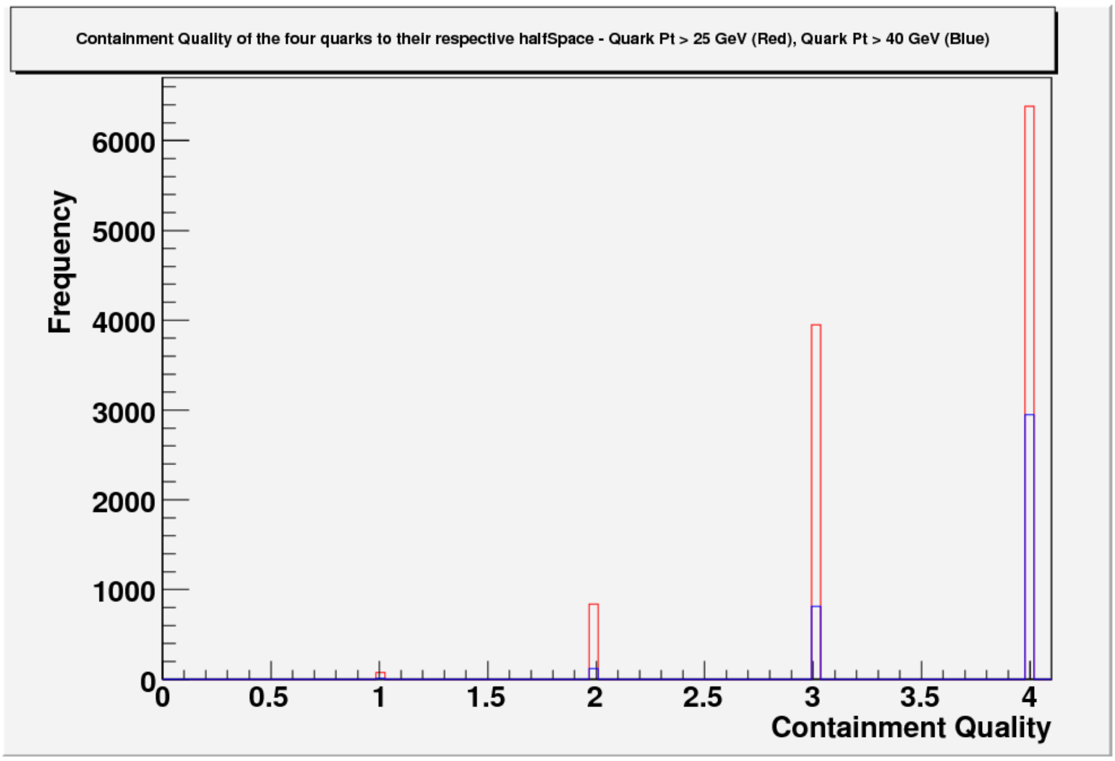

Figure 17: Containment Quality of four quarks to their respective half-space, as defined by top particles. Quark p_T greater than 25 GeV (red) and P_t greater than 40 GeV (blue)

Figure 17 shows the containment quality for the four quarks to their correct half-space, with the half-space normal being defined by the top quark momentum vectors. Plots have been made for two quark p_T cuts, a soft cut of 25 GeV (red) and a hard cut of 40 GeV (blue). From the plots, with the cut of 25 GeV, it can be seen that many events are have a containment quality of four, around 6500 events. The number of events with values less than this then fall of rapidly, with only approximately 4000 events with a value of three (one quark having escaped its half-space) and only approximately 800 events with a value of two (two quarks having escaped their half-space). There are then negligible events with a containment quality of one, and zero events with a quality of zero. These results are promising as this indicates that many of the events at the generated quark level can be considered 'perfect' with respect to the containment of the four quarks. In addition, there are many more events with higher values of containment quality than those with lower values.

Moving to the hard momentum cut of 40 GeV, the results are improved further. There are now considerably more events with containment quality of four than any other value, and the number of events with a value of two are negligible. However, it can be seen from comparing the red and blue distributions that the hard momentum cut has dramatically reduced the number of events in the sample for the hard momentum cut.

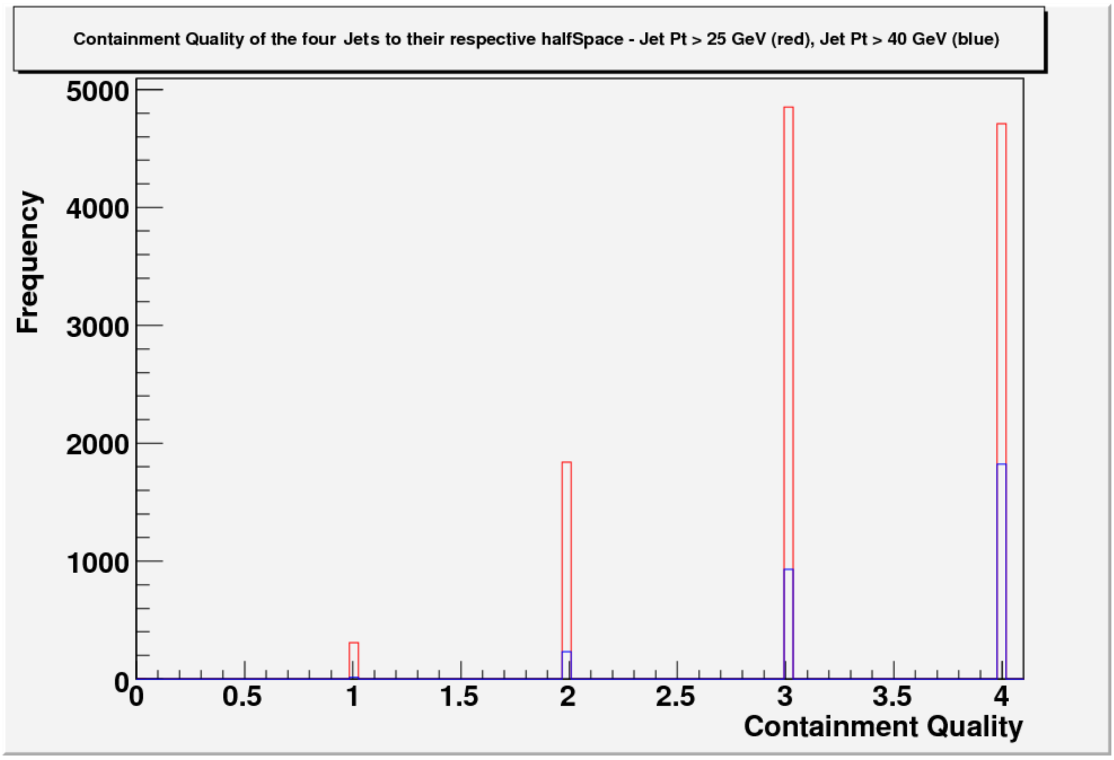

Figure 18: Containment Quality of four jets to their respective half-space, as defined by W particles. Jet p_T greater than 25 GeV (red) and P_t greater than 40 GeV (blue)

Figure 18 shows the containment quality for the four jets to their correct half-space, with the half-space normal being defined by the W particle momentum vectors. Plots have been made for two jet p_T cuts, a soft cut of 25 GeV (red) and a hard cut of 40 GeV (blue). The situation for this histogram is quite different to that of figure 17. The quarks have been replaced by jets, implying the problems discussed earlier. In addition, the half-space plane is now defined using the W momentum vectors, as these are much easier to reconstruct in the detector than the top quark momentum vectors.

As of these changes, the results are quite different, although still promising. With the soft cut of 25 GeV, it can be seen that there are roughly equal numbers of events with values of both three and four. Again, there are fewer events with a value of two, and negligible amounts with one or zero. The quality of the events in this plot are much worse than those in figure 17, mostly due to the definition of the half-space normal with the W particle vectors. However, the quality of the results are regained with the hard p_T cut of 40 GeV, where there are approximately twice as many results with containment quality of four than three, and relatively small amounts with a value of two. Again, there are only negligible events with values of one or zero.

3 Summary

The results gathered from the investigation have been positive. Both the topics investigated, ΔR distributions for correct and incorrect combinations of jets, and the half-space studies splitting the event into two separate decay chains, have gained results that can be used as input for a likelihood function. In the case of the ΔR investigations, it is the ability to distinguish the distributions of correct and incorrect jet combinations, for both jet p_T cuts made, that makes the quantity a valid input. For the half-space studies, it is the relative abundance of events with containment quality value of three or four (three or four jets contained to their correct half-space respectively), for both jet p_T cuts.

The investigation can also be extended to potentially achieve better results. For example, cut optimisation will immediately change the distributions in both the ΔR and half-space investigations. For the half-space investigations particularly, investigating if better results can be obtained by alternative definitions of the half-space,for example, not using the momentum unit vector of the two particles to define the plane may improve results in cases where the particle momenta differ. In addition, no study into having multiple half-space planes, or defining planes between other particles in the decay have been done. Also, analysis of which particles are likely to leave their half-space, and under what conditions, could improve the results farther.