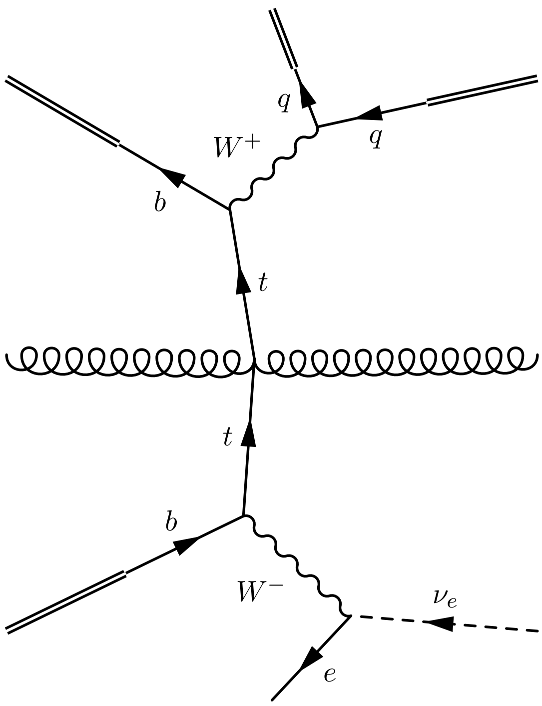
Evaluation of Jet Reconstruction Algorithms for a Measurement of the Top-Quark Mass in the $t\bar{t} \to \text{lepton} + \text{jets}$ Channel at ATLAS

Patrick Rall
Candidate Nr.: 000169065
Session: May 2012
Munich International School

IB Extended Essay in Physics
Supervisor: Mr. David Pritchard

Word count: 3993

November 14, 2011



Abstract

In order to perform a mass determination of the top-quark, a so-called $t\bar{t}$-event is analyzed. The $t\bar{t}$-event is a type of event that can occur at the ATLAS experiment at the Large Hadron Collider (LHC) at CERN, which produces two top-quarks that rapidly decay into further decay products, which are finally visible as jets inside the calorimeters[1]. For the mass determination, in order to reconstruct the original top-quarks, these jets need to be correctly identified and attributed to certain processes using a jet reconstruction algorithm.

In this analysis three jet reconstruction algorithms are investigated and evaluated in terms of reconstruction efficiency and suitability for the purposes of the mass determination. The identification of a specific pair of 'light jets' produced by a W boson decay in the $t\bar{t}$-event proves to be a very important concern, as an additional W boson reconstruction is necessary for the reduction of systematic uncertainty.

The assumptions made by a 'p_T-max' method and a 'ΔR' method prove to be very effective in top mass reconstruction, but are limited in effectiveness concerning light jet identification. A 'm_W-match' method proposed in 2011 is specifically designed to be effective in W boson reconstruction and thus is significantly more effective than the other methods, although still leaving room for improvement.

Furthermore, a set of angle cuts based on the assumption of the 'ΔR' method is evaluated. These are shown to significantly enhance jet reconstruction efficiency. However, only some of the angle cuts applied appear to be useful to the performance. The use of the angle cuts is recommended, but the methods involved require refinement.

It is hence determined that the 'm_W-match' method together with the angle cuts is the most suitable jet reconstruction algorithm of those considered.

287 words.

[1] In particle physics a calorimeter is a measuring device that can measure the energy of particles. There exist various types and designs with different purposes within the ATLAS experiment.

Contents

Acknowledgements **3**

1 Introduction **4**

2 Method used for measurement of the top-quark mass **4**
2.1 $t\bar{t}$-events 4
2.2 Jet reconstruction procedure and mass computation 6
2.3 Parameterized curve fits on monte-carlo data 6

3 Approach for analysis of algorithms **7**
3.1 Summary of methods 7
3.2 Criteria of analysis 8
3.3 Procedure of analysis 9

4 Evaluation of jet-reconstruction algorithms **9**
4.1 'p_T-max' method 9
4.2 'ΔR' method 11
4.3 'm_{W}-match' method 12
4.4 Applied angle cuts 13

5 Conclusion **17**

References **18**

Acknowledgements

I express my gratitude to:

David Pritchard, Director of Instructional Technology at Munich International School, for providing me with helpful comments and assisting me in the planning and structuring of this analysis.

Dr. Richard Nisius, ATLAS-Inner Detector Group Leader at the Max Planck Institute for Physics, for granting me the opportunity to contribute to the research on the top-quark mass, teaching me large amounts of technical and theoretical knowledge which was critical to the making of this analysis, and reviewing my work.

Dr. Giorgio Cortiana, member of the ATLAS-Inner Detector Group at the Max Planck Institute for Physics, for answering questions concerning the implementation of the 2010 jet reconstruction algorithms, the ROOT[2] data analysis framework, and giving me access to Monte-Carlo simulation data.

Dr. Nelia Mann, instructor for the 'Particle Physics' course at Stanford EPGY[3] Summer Institutes, for giving me a great general education in particle physics during Stanford EPGY Summer Institutes 2009.

[2]A C++ library commonly used by particle physics researchers for data processing and visualization.
[3]Education Program for Gifted Youth at Stanford University, CA.

1 Introduction

At the LHC at CERN, protons are accelerated to very high energies and collided with each other inside particle detectors like the ATLAS detector. These energetic collisions may produce interesting quantum mechanical interactions, including the production of the top-quark. As quarks are highly unstable, they quickly form decay products in the form of a jet, which are collected inside calorimeters in the detector. Many interesting analyses can be made by studying the jets, including top-quark mass determinations.

Jet reconstruction algorithms are an essential component of many analyses in particle physics as they constitute the only connection between the quantum mechanical interactions of quarks and the data collected by particle detectors.

The events inside the detector often produce jets resulting from quarks produced by quantum mechanical interactions. Jets can be added together in order to compute the mass of the original quarks. This is extremely useful for any hadronic decay into quarks, rather than into leptons. However, to perform this effectively the various jets emitted in the event must be correctly identified using directional and energy information only. Due to this limitation, the identification process requires the application of complicated techniques which work on various assumptions. Various different algorithms exist, each with a different degree of effectiveness.

In this paper, three jet reconstruction algorithms are investigated for a top-quark mass determination. The methods use different properties and assumptions about the events of interest at the ATLAS detector, the $t\bar{t}$-events, in which the top-quark is produced, and use different techniques to identify jets. These algorithms are evaluated for quality taking into consideration the top-quark mass determination methodology.

Hence, my research question is: Which jet reconstruction algorithm is the most suitable for this top-quark mass measurement procedure?

2 Method used for measurement of the top-quark mass

In [1], the prospects for a measurement of the top-quark mass m_{top} are presented. The procedure of this mass determination will be the focus of this analysis. Jet reconstruction algorithms do not just represent a critical component of this process, but must be analyzed in a fashion that takes the rest of the process into consideration. Here, the measurement process of [1] is summarized and points that must be taken into consideration are described.

2.1 $t\bar{t}$-events

To find the mass the top-quark, it must first be produced in a particle interaction. Unfortunately, quarks have very short lifetimes, making them impossible to detect directly. Instead, they decay into various other particles which in turn decay into more particles, and so on. Eventually, these are visible as a jet of particles when measured by the detector. These measurements can then be reconstructed to gather information about the original jets, and finally the original quarks.

The top-quark mostly appears in a process called a 'tt̄-event'. This event will be the focus of this analysis. A top-quark t and its antiparticle, the anti top-quark t̄ are produced mostly in a gluon-gluon interactions. The most likely decay path for these particles is a bottom-quark b and a W boson [2]. The W boson further decays into either two lighter quarks q or a lepton and its corresponding anti-neutrino. This analysis will focus on events where one of the W bosons decays leptonically (i.e. into a lepton and neutrino) and the other into quarks. The detector finally measures a total of four jets from the light quarks and b-quarks, the lepton and a missing transverse energy[4] from the neutrino. To further narrow the selection only the electron channel[5], that is, where the W boson decays into an electron e and an electron anti-neutrino $\bar{\nu_e}$ is considered. See Figure 1.

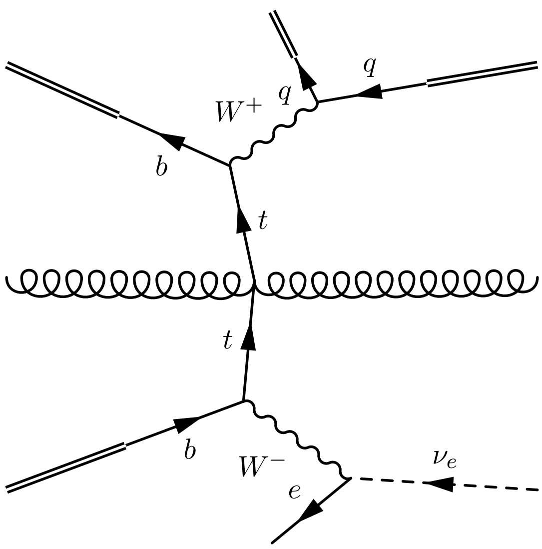


Figure 1: Example of a Feynman diagram-like schematic of a tt̄-event of the variety that will be considered in this analysis. The direction of the lines indicates particles and antiparticles. Doubled lines represent jets produced by decaying quarks. Curly lines represent gluons, wavy lines W bosons, and dashed lines the invisible neutrinos. The event where the W^+ decays leptonically and the W^- hadronically is also considered.

[4]Being particles that only perform the extremely rare weak interactions, neutrinos are impossible to measure inside the ATLAS detector. However, they can be identified by the fact that a component will be missing from the otherwise zero transverse energy, calculated from the other measurements.

[5]Data from particle detectors are divided into channels, which are simply data resulting from different processes. One can refer to a tt̄-event channel containing all tt̄-events, or more specific channels like the electron or muon channels for the tt̄-event.

2.2 Jet reconstruction procedure and mass computation

The ATLAS detector's measurements are combined into jets using a cone algorithm[6] [2]. This provides easy access to reconstructed jet information. Since a wide variety of events occur at the LHC, various event selections are made to the event data based on this information in an attempt to remove as many background (i.e. non-$t\bar{t}$-events) events as possible. These selections are identical for all reconstruction algorithms later to be analyzed, and are described in detail in [3]. They include measures to ensure the presence of exactly one electron, one neutrino, a sufficient number of jets, as well as the proper isolation of jets from each other. A selection criterion that will be omitted from this analysis is the W boson mass window, which removes events where the computed W boson mass is outside a certain range. The reason for this is that the W boson mass reconstruction efficiency is a criterion for the effectiveness of the jet reconstruction algorithms.

Once an event has passed initial event selection it is treated as a $t\bar{t}$-event, even though some background events may still be present. The jet reconstruction algorithms then use the information available to determine which of the four jets present corresponds to the light quarks emitted by the W boson and the two b-quark jets. This way, the reconstructed mass of the top-quark $m_{\text{top}}^{\text{reco}}$ and the reconstructed mass of the W $m_{\text{W}}^{\text{reco}}$ can be computed for an event.

In a calorimeter, not all incident energy is collected. The Jet Energy Scale (JES) is a process that scales the energy of each jet from the value picked up in the calorimeter to the expected input energy. Much of the systematic uncertainty of this analysis originates from the uncertainty of the JES. In order to mitigate this, a new stabilized top mass $m_{\text{top}}^{\text{stab}}$ is computed from the measured values. By computing a ratio between the two reconstructed quantities to which the JES is applied, $m_{\text{top}}^{\text{reco}}$ and $m_{\text{W}}^{\text{reco}}$, the effects of the JES uncertainty should largely cancel. In order to bring the ratio back to a value similar that of m_{top}, it is multiplied by a literature value of the W boson mass.

$$m_{\text{top}}^{\text{stab}} = \frac{m_{\text{top}}^{\text{reco}}}{m_{\text{W}}^{\text{reco}}} m_{\text{W}}$$

Although this technique has been shown to eliminate much of the systematic uncertainty [1], it requires the correct identification of the light jets by the reconstruction methods in order to compute $m_{\text{W}}^{\text{reco}}$ in addition to just $m_{\text{top}}^{\text{reco}}$.

2.3 Parameterized curve fits on monte-carlo data

In order to perform a measurement of the top-quark mass, the individual jets must be reconstructed from the particles in the detector, identified, and properly combined. However, the result of performing this procedure will yield a distribution of masses,

[6] A cone algorithm, as opposed to a clustering algorithm, attempts to place a conical shape around each jet and adds up everything within it. Clustering algorithms, like the anti-k_t algorithm, instead focus on the addition of component particles that are close to each other until a certain stopping point is reached.

rather than a single value, partly due to the probabilistic nature of quantum mechanics but mostly due to the measurement uncertainties. The invariant mass of the top-quark, however, is a single value, namely the peak of the probability density function of the measured top-quark masses. To find this value, the analysis is first performed on data generated by a Monte-Carlo procedure[7], where various values of m_{top} can be entered. Then, parameterized curve fits are performed, and the fit parameters matched to the specified value of the top-quark mass. Finally, the curve fit is performed on data collected at ATLAS, and the real-world m_{top} is computed.

This has two consequences on jet reconstruction algorithms. First, it provides an environment suitable for testing an algorithm's efficiency, as the actual identities of the jets are known in the format of so-called truth jets. Secondly, these parameterized curve fits partly compensate for background events, as the background elements of the distribution are identical for different values of m_{top} and hence will not affect the fit parameters much.

Since the focus of this analysis is jet reconstruction, the parameterized curve fits affect the criteria but will not be performed on the resulting distributions.

3 Approach for analysis of algorithms

This section presents the method used for the analysis of the individual jet reconstruction algorithm and compares the results and establishes a set of criteria.

3.1 Summary of methods

In this investigation the three jet reconstruction methods will be analyzed. The first two were proposed in 2010 [1]. In 2011, another algorithm that uses a slightly more sophisticated method was proposed, which makes use of some of the developments since the last year, including larger amounts of data available from ATLAS.

The 'p_T-max' Method identifies the top-vector by finding the permutation of three jets that maximizes transverse momentum p_T. This is expected to discard the b-jet emitted by the leptonic top-quark, as it usually travels in the opposite direction. Then, the two jets with closest angle in between them will be assumed to be the light jets emitted by the W boson. This process works best if the top quarks are emitted perpendicularly to the central axis of the detector. If it is not so, then transverse momentum is not fully indicative of direction.

The 'ΔR' Method makes the simple assumption that the two jets with the highest p_T will be the b-jets, and the two remaining jets will be the light-jets. The b-jet of the hadronic top is assumed to be closer to the reconstructed W-vector than that of the leptonic top. It is heavily reliant on angle cuts, which are described below.

[7] Monte-Carlo procedures are simulations of processes that are well explained in theory, and utilize random numbers to generate results. Here, they create data that are very similar to those from the ATLAS detector, but of course here the mass of the top-quark can be specified in the simulation.

The 'm_W-match' Method is a method that uses the fact that we have very precise measurements of m_{W} available. It computes the pair of jets with the highest probability of being the two light jets by applying a scale function to each of the jets combined and comparing the resultant mass to that of the W boson. A constraint is applied to the scale function based on the width of a jet distribution. These factors are integrated into a χ^2 formulation[8]. Then, the b-jet is identified using a method similar to the p_T-max method.

Since the scale function resembles the purpose of the Jet Energy Scale (JES), computation of $m_{\mathrm{top}}^{\mathrm{stab}}$ makes little sense here. As the values of the jets from the scale function are expected to be more representative of the actual jet energies than those from the JES, the computation of the reconstructed masses should utilize the computed scale factors. In this analysis however, it is the combinatorial efficiency that is of interest, and thus the m_W-match method's χ^2 minimization will be used for jet identification only.

Angle cuts are an additional element of the ΔR method that can also be applied to other methods. They eliminate events based on the light jets being too far apart or the b-jets being too close together, using the same assumption-based identification process as in ΔR method.

3.2 Criteria of analysis

The most significant criterion for the evaluation of algorithms is how many events are correctly reconstructed. Depending on exactly what jets are misidentified, there will be false values of the stabilized top mass $m_{\mathrm{top}}^{\mathrm{stab}}$ that appear close to the signal region. If the number of incorrect events is large, the perturbations will be significant in the parameterized curve fits. The difficulty of avoiding this is enhanced by the fact that in order to compute $m_{\mathrm{top}}^{\mathrm{stab}}$, both $m_{\mathrm{top}}^{\mathrm{reco}}$ and $m_{\mathrm{W}}^{\mathrm{reco}}$ must be correct.

Since events within sidebands, i.e. events outside the signal region, may originate from both incorrectly reconstructed events and background events, the parameterized curve fits will be to some extent capable of compensating for uncertainties causing a large change in the measured value. Consequently, the most significant incorrect reconstructions are those in the signal region, i.e. with mass values close to m_{top}. Incorrect sideband reconstructions can largely be ignored, so long as they are not so numerous that they shift the position of the mean.

Performance in W mass reconstruction is not only important in the computation of $m_{\mathrm{top}}^{\mathrm{stab}}$, but also as a means of identifying background events. This makes this criterion important for the m_W-match method as well, which does not necessitate $m_{\mathrm{top}}^{\mathrm{stab}}$, as in practice it uses the scale factors computed by the χ^2 minimization directly.

A final consideration is that not too many signal events are eliminated by any cutting procedures, as these may result in a bias.

[8] A χ^2 minimization is a statistical technique used commonly for fitting parameters to a data model. A χ^2 formulation expresses the deviance of the fit from the data in a Gaussian fashion. This χ^2 term is minimized for the values of the parameters which produce the best fit.

3.3 Procedure of analysis

In order to compare the effectiveness of jet reconstruction methods, the degree to which they fulfill the criteria must be expressible in a concrete numerical metric. What is formulated here is the reconstruction method's 'efficiency': the percentage of events in the signal region where jets were identified correctly.

Methods are applied to about 1 200 000 events from a Monte Carlo simulation using various algorithms as described in [1], containing only $t\bar{t}$-events. Particles detected in the simulated environment will be reconstructed into jets using a cone algorithm. The result will be an array of jets represented by four-vectors[9]. This data format is similar to that provided by ATLAS. In addition to detectable data, however, the simulation also yields correct matched jets for each of the quarks emitted by the $t\bar{t}$-events.

First, the event selection procedure, which is identical for all algorithms will be applied to all event data. The requirements are as described in section 2.2. Then angle cuts are applied. The W and top vector are reconstructed using the respective algorithm. Matched jets are also reconstructed accordingly. The resulting match W and top vectors are compared to the reconstructed vectors. If the angle between the vectors is less than 0.2 radians, and 80% of matched jet energies are less than the reconstructed jet energies, then the reconstruction is considered correct [2]. Reconstructed jet masses are filled into separate $m_{\mathrm{W}}^{\mathrm{reco}}$ and $m_{\mathrm{top}}^{\mathrm{reco}}$ histograms.

Reconstruction efficiency is judged using analysis of the histograms, as well as a metric for efficiency. Since correctly identified events are very close to the mass in question, a Gaussian fit is applied to the histogram of correctly identified events. A mass region of $\pm 3\sigma$ is considered to primarily consist of signal data. Inside the signal region, the ratio of correctly identified events N_i over total events N_t is used to compute a % efficiency. Error δ of this quantity is computed according to [2]:

$$\delta = \sqrt{\frac{\frac{N_i}{N_t}(1 - \frac{N_i}{N_t})}{N_t}}$$

4 Evaluation of jet-reconstruction algorithms

This section addresses the advantages and disadvantages of each of the jet reconstruction algorithms in turn, and then investigates angle cuts. For the initial analyses, 2.5 radians is the maximum angle between light jets and also the minimum angle between b-jets. These values were chosen in accordance to the recommendation in [3].

4.1 'p_T-max' method

The p_T-max method first identifies the top-quark by finding the sum of three jets out of the four jets total that maximizes transverse momentum, essentially removing the b-jet

[9] Lorentz vectors or four-vectors represent vectors that describe both directional and energetic information. A description of the coordinate system is available in [1].

from the leptonic top. The light jets are taken as the pair of jets with the smallest angle between them within the jets selected by transverse momentum maximization.

Results of a simulation with all events is shown in Figure 2. Entries are scaled to an integrated luminosity[10] of 100 pb^{-1}, as are all later histograms.

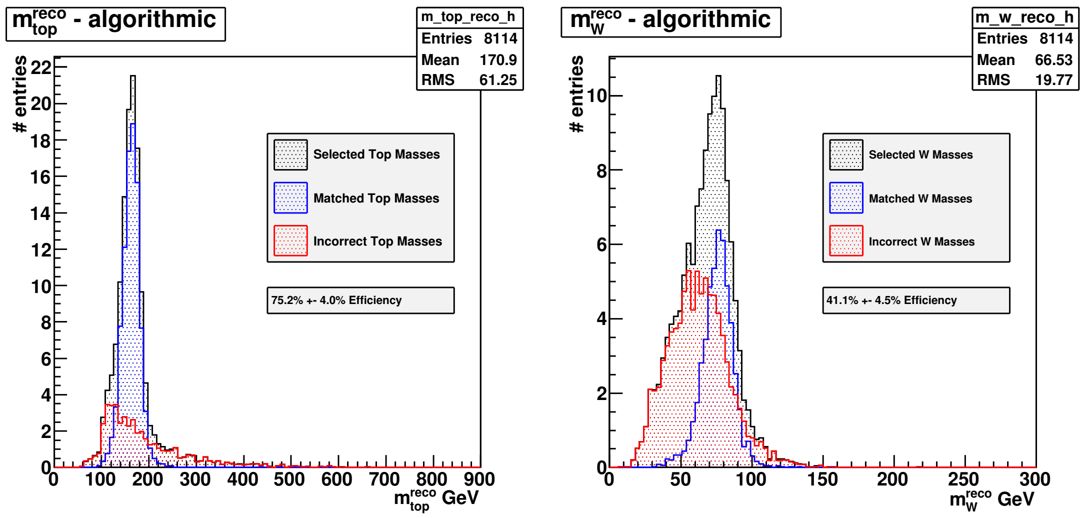


Figure 2: Histograms of p_T-max method reconstructions of $m_{\mathrm{top}}^{\mathrm{reco}}$ and $m_{\mathrm{W}}^{\mathrm{reco}}$. Events with correct top and W vector reconstructions are shown in blue, incorrect reconstructions in red. Total reconstructions are in black, and are the sum of the correct and incorrect reconstructions. Cut events are not shown. This scheme extends to all following histograms. 8114 events passed the event selection and angle cuts. However, due to luminosity scaling, the number of entries in the histogram is fewer.

A few things are evident from these histograms. First, the $m_{\mathrm{top}}^{\mathrm{reco}}$ reconstructions appear to be very satisfactory, with a $75.2\% \pm 4.0\%$ efficiency, and the vast majority of the peak region correctly reconstructed. This makes sense, as for $m_{\mathrm{top}}^{\mathrm{reco}}$ the operation requires nothing more than the correct elimination of one jet: the b-jet from the leptonic top. Not only is the number of incorrect reconstructions small, but many of these are also in the sideband region which is negligible. Incorrect reconstructions that are closer to the peak at ∼120 GeV are a greater concern. Since previous measurements of m_{top} yield ∼173 GeV [1], this may result in a bias. However far from perfection the top-quark reconstruction is, the effects are not very severe.

Concerning $m_{\mathrm{W}}^{\mathrm{reco}}$, the number of incorrectly identified events is striking. Not only is the number of incorrectly reconstructed events in the signal region greater than that of correct reconstructions, but the peak of incorrect reconstructions at ∼60 GeV makes even the presence of a 'signal region' as defined in Section 3.3 questionable. Histogram mean is 66.53 GeV, far off from $m_W = 80.426$ GeV [4]. Since correct reconstruction of $m_{\mathrm{W}}^{\mathrm{reco}}$ is essential for computation of $m_{\mathrm{top}}^{\mathrm{stab}}$, this result is extremely unsatisfactory, as a

[10]Integrated luminosity is the number of particles delivered by a particle accelerator per unit area integrated with respect to time. Picobarn (pb) are a unit of area.

large proportion of the correct $m_{\rm top}^{\rm reco}$ reconstructions will be scaled incorrectly.

Conclusions to be drawn from this is that the p_T maximization, which is only used for the computation of $m_{\rm top}^{\rm reco}$, is highly effective for this purpose. However, the method of identifying the light jets must be strongly refined.

4.2 'ΔR' method

The ΔR method sorts the four jets by p_T, and assumes the two with highest p_T to be the two b-jets, and the lowest two to be the light jets. Then, the b-jet with the smallest angle to the W vector is taken as the one from the hadronic top. Note that, contrary to the p_T-max method, this technique finds the W vector first, unlike the inefficient $m_{\rm W}^{\rm reco}$ reconstructions of the p_T-max method. Histograms from identical data as in Figure 2 are shown in Figure 3, here using the ΔR method.

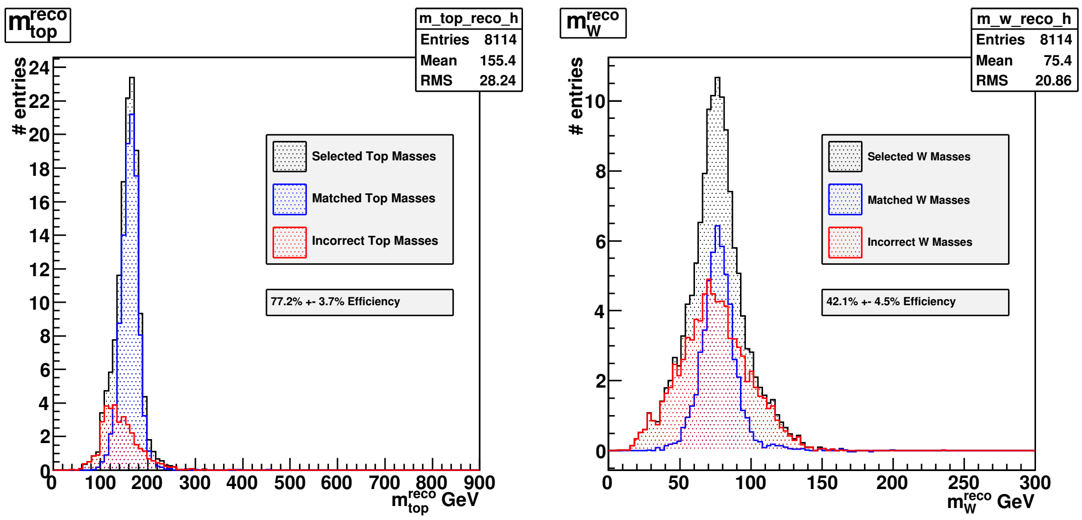


Figure 3: Histograms of ΔR method reconstructions of $m_{\rm top}^{\rm reco}$ and $m_{\rm W}^{\rm reco}$. Negative event magnitudes originate from the Monte-Carlo program's means of dealing with negative terms in higher-order Feynman diagram evaluations: events are assigned a weighting that may be negative in some cases. With large amounts of data, these negative totals should never appear. However, in sideband regions where there are very few events, these can result in negative event counts.

Performance in computation of $m_{\rm top}^{\rm reco}$ of the ΔR method is comparable with that of p_T-max, them being the same within uncertainty. However, the mean of the $m_{\rm top}^{\rm reco}$ histogram in Figure 3 is 155.4 GeV, further away from $m_{top} \approx 173$ GeV [1] than that of the p_T-max technique where the mean was 170.9 GeV. This is due to a similar peak structure of incorrect reconstructions as in Figure 2, just more pronounced as it lacks the additional high-mass sideband reconstructions.

Effectiveness in W boson reconstruction continues to be very low. In this context however, the failure of identification implies that the assumption fundamental to the ΔR method, namely that the jets with lowest p_T are the light jets, is correct less than half (42.1% ± 4.5%) the time. The effects of this have a fairly high chance of being corrected in the computation of $m_{\text{top}}^{\text{reco}}$: Since it is not likely that both b-jets have less transverse momentum than the light jets, incorrect light jet identifications will usually include one light jet and one b-jet. Especially if the b-jet was from the hadronic top-quark, the jet selected for closest angle to the W vector will be the other light jet, resulting in a correct top reconstruction from a false identification. Regardless of the lack of reliability in the central ΔR assumption, it still outperforms the p_T-max method in W reconstruction slightly, not because of a higher efficiency, but rather because of a mean $m_{\text{W}}^{\text{reco}} = 75.4$ GeV closer to the correct $m_W = 80.426$ GeV [4], as the distribution of incorrect events is closer to the center.

As before, the main source of concern is the reconstruction of the W boson. Neither the p_T-max nor ΔR techniques can identify it with a sufficient reliability. With the ΔR method however, the effectiveness of the main assumption in the computation of $m_{\text{top}}^{\text{reco}}$ originates only from the high probability of its misidentifications being corrected by chance. This must be taken into account when evaluating angle cuts, which are based on this same assumption.

4.3 '$m_{\mathbf{W}}$-match' method

The m_{W}-match method addresses the issue of light jet identification most directly. For each permutation of two jets in the four, it performs a χ^2 minimization using the TMinuit library [5]. In order to do this, it applies two scale factors to the light jets in question, denoted by α_1 and α_2. TMinuit adjusts these factors to minimize χ^2. Then the jet pair with minimum χ^2 is selected for $m_{\text{W}}^{\text{reco}}$.

The χ^2 formula takes into account two criteria. It has been shown that the W boson appears as a Breit-Wigner distribution[11] when measured, with peak mass $m_W = 80.426$ GeV and a distribution width $\Gamma_W = 2.150$ GeV [4]. In this analysis this is approximated by the Gaussian distribution that χ^2 creates, with the parameters of the Breit-Wigner distribution, hence the first term of the formula. The second two terms are used to prevent the scale factors from simply adjusting the jets to perfection, by comparing the scaled jet energies to the original energy of the jets. Distribution of jet energies of the individual jets is also assumed to be Gaussian, with a width assumed to be 50% of the jet energy, which roughly agrees with the observed width [2]. With the scale factors denoted by α_1 and α_2 and the jet four-vectors by $\vec{j}_1$ and $\vec{j}_2$, and $M(\vec{x})$ and $E(\vec{x})$ the mass and energy respectively, the formula is:

$$\chi^2 = \left(\frac{M(\alpha_1\vec{j}_1 + \alpha_2\vec{j}_2) - m_W}{\Gamma_W}\right)^2 + \left(\frac{E(\vec{j}_1) - E(\alpha_1\vec{j}_1)}{\frac{1}{2}E(\vec{j}_1)}\right)^2 + \left(\frac{E(\vec{j}_2) - E(\alpha_2\vec{j}_2)}{\frac{1}{2}E(\vec{j}_2)}\right)^2$$

Results of a simulation with all events is shown in Figure 4. Although all reconstructed masses collect around the peak, making the mean 81.85 GeV, as would be

[11] A Breit-Wigner distribution is a probability density function similar, but not the same in shape to a Gaussian distribution. It is often used in particle physics.

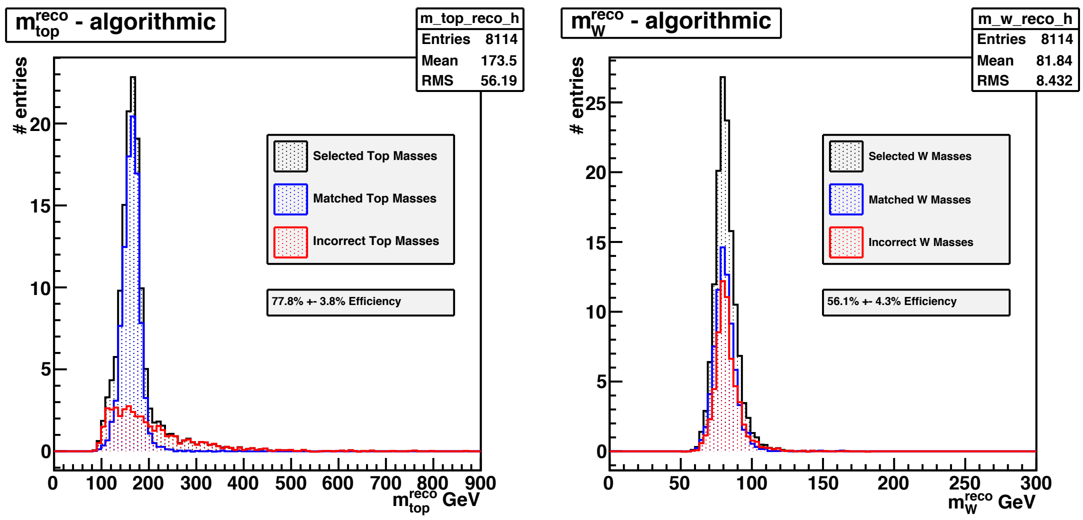


Figure 4: Histograms of m_{W}-match method reconstructions of $m_{\mathrm{top}}^{\mathrm{reco}}$ and $m_{\mathrm{W}}^{\mathrm{reco}}$.

expected an algorithm of this nature, still a significant fraction of them are incorrect: efficiency is $56.1\% \pm 4.3\%$. Since the nature of the distribution can be compensated for by parameterized curve fits only in top mass distributions, it is mainly this raw fraction of efficiency that is of interest. However, the m_{W}-match method is a significant improvement over the other algorithms in W boson reconstruction, but still contributes significant systematic uncertainties. $m_{\mathrm{top}}^{\mathrm{reco}}$ computation performance is comparable to the other two algorithms.

4.4 Applied angle cuts

The angle cuts were originally a component of the ΔR method, and use the same assumption about the ordering of jet p_T to identify the b and light jets. They can easily be applied to other algorithms. Although the inaccuracy of the main assumption was shown in section 4.2, they have still proven to significantly improve efficiency, see Figure 5. Here it is visible that the largest benefits are for the top reconstruction.

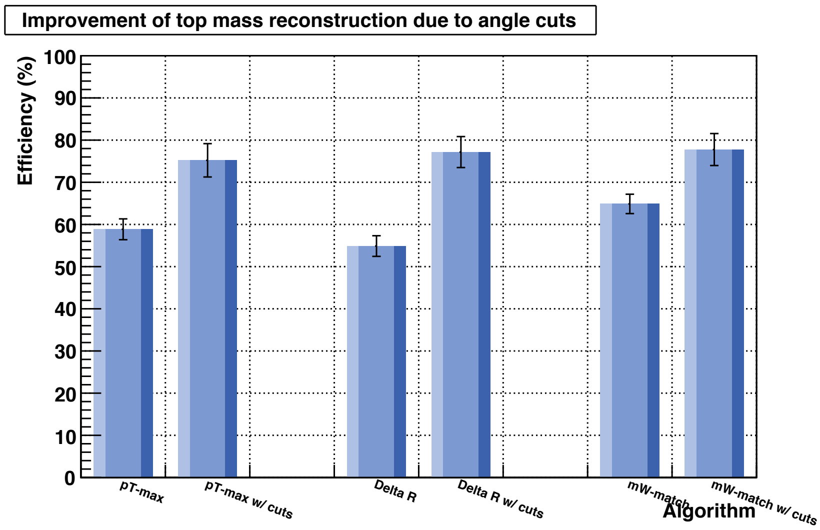
This benefit in efficiency is very desirable. However, if the cuts affect signal data in a fashion that may create a bias in the mass determination they still should not be included. In order to test this, cut events have been collected into an additional histogram superimposed on output from a ΔR method based reconstruction, so that the effects of the cut can be visualized. This is done separately for the light jet and b-jet conditions. See Figures 6 and 7.

In Figure 6 it is visible that cuts occur in a distribution similar to that of the non-cut events, implying that it is unlikely that a bias is caused by this procedure. Some additional events are cut in the sidebands, which is also a benefit. However, with the amount of data available, this is hardly a problem. Already this procedure alone has increased the top reconstruction efficiency from $54.9\% \pm 2.4\%$ to $73.8\% \pm 3.1\%$ and the

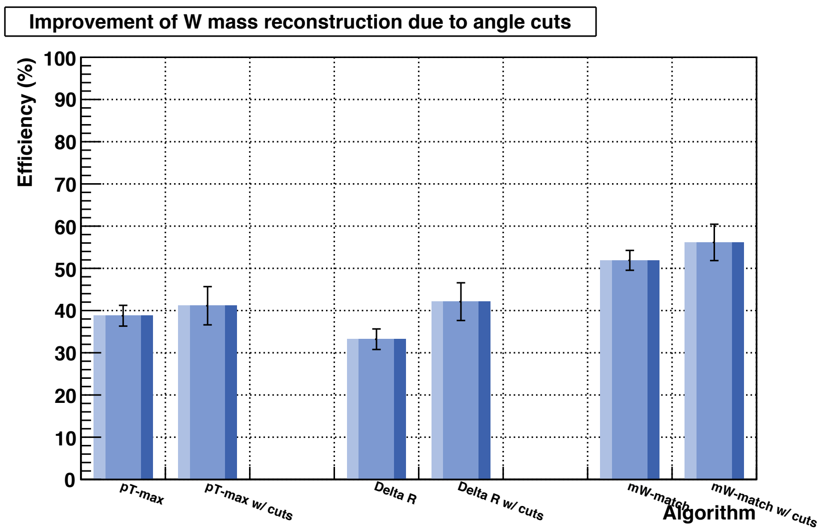
W reconstruction efficiency from $33.2\% \pm 2.4\%$ to $41.8\% \pm 3.7\%$.

In Figure 7, which displays output with the condition of a maximum angle between light jets, one can see that in both cases the majority of cut events are located at one side of the overall distribution. This is a clear example of bias, as the cut events move the mean of the reconstructed masses to the side of higher mass, especially with the W reconstructions. Also, the efficiency improvement is absolutely minimal and within uncertainty of reconstruction efficiencies without this condition. The reason for this lack of performance is due to the inability of the ΔR method assumption to identify light jets reliably.

This analysis reveals that the main source of improvement due to angle cuts is due to the condition of a minimum angle between *b*-jets. It is suggested that the angle cut on light jets be revised in terms of cutoff angle and identification method, or omitted completely.



(a) Top reconstruction



(b) W reconstruction

Figure 5: Comparison of efficiency improvement due to angle cuts for different jet reconstruction algorithms.

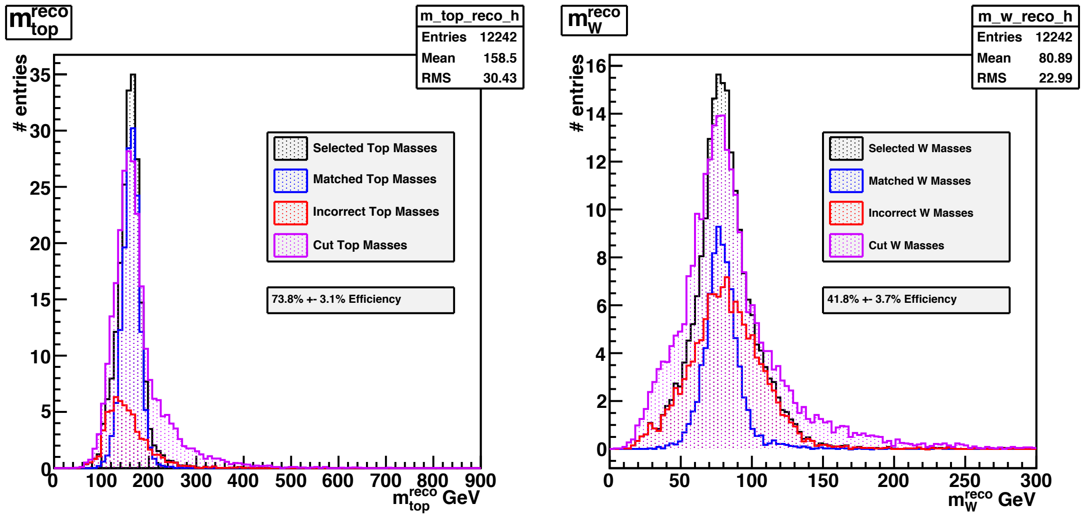


Figure 6: A ΔR method reconstruction showing events cut by a condition of a minimum angle of 2.5 radians between b-jets in addition to the masses and combinatorial efficiency of non-cut events.

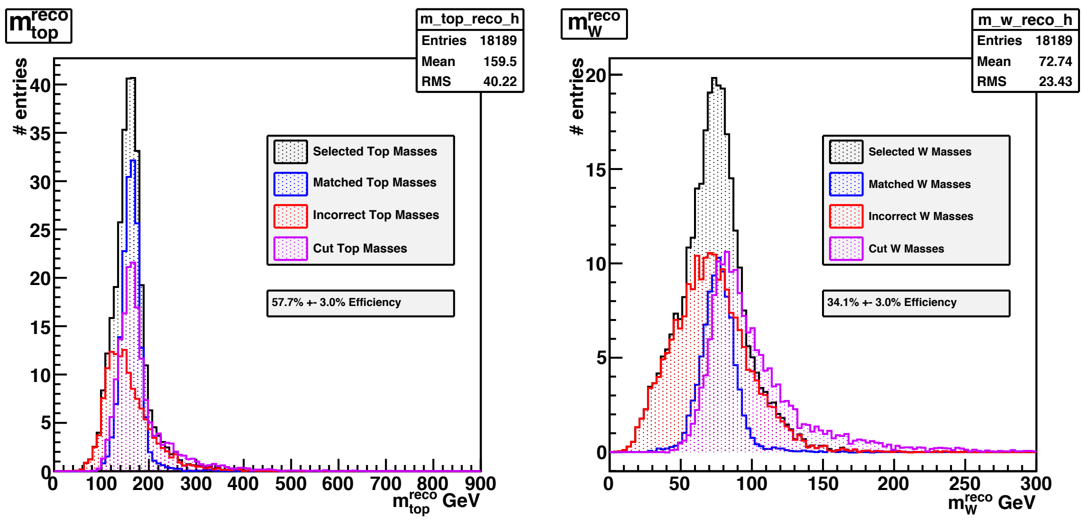


Figure 7: Diagram same as in Figure 6, but with a cut condition of a maximum angle of 2.5 radians between light jets instead of a cut on the angle between b-jets.

5 Conclusion

In the analysis of the three jet reconstruction algorithms, all methods showed advantages in some areas. The p_T-max method proved to have a very reliable assumption that the b-jet from the leptonic top can be eliminated by choosing the set of jets that maximized p_T. The ΔR method was capable of achieving a similar performance in top reconstruction with a much simpler assumption, that showed to create less bias in W boson reconstruction. However, the main problem that the approaches faced was that of correct identification of light jets, a task that is critical for the computation of $m_{\text{top}}^{\text{stab}}$. Even in the m_W-match algorithm, which does not make use of the $m_{\text{top}}^{\text{stab}}$, W boson reconstruction is very useful for the elimination of background events.

The m_W-match algorithm is the most effective in the computation of $m_{\text{W}}^{\text{reco}}$, while bringing equal efficiency in that of $m_{\text{top}}^{\text{reco}}$, namely one of $77.8\% \pm 3.8\%$, with minimal bias. Even so, with only $56\% \pm 4.3\%$ efficiency in W boson reconstruction, the technique applied in the method should be refined, possibly using a likelihood technique[12].

The angle cuts are shown to bring significant improvement to reconstruction efficiencies. However, the b-jet cuts are the main source of this improvement and the light jet cuts make little contribution and even add bias. This is attributed to the inefficiency of the ΔR method's central assumption. It is thus recommended that the application of the b-jet cuts alone be considered, or, more preferably, that a different assumption than that of ΔR is used for angle cut jet identification.

Disregarding the further analyses to be made on improvements on statistical techniques and angle cuts, the m_W-match method with use of angle cuts gives the highest performance in jet reconstruction and is hence the most suitable algorithm of the ones investigated for this top-quark mass determination.

[12] Maximization of a likelihood function is an advanced statistical technique that is not constrained to a Gaussian probability distribution like the χ^2 minimization, and would thus be capable of integrating the precise shape of the Breit-Wigner and jet distributions.

References

[1] G. Cortiana and R. Nisius, "Measurement of the top-quark mass using the stabilized mass variable via the template method in tt-bar to lepton+jets channel at ATLAS," *ATLAS Note*, vol. ATL-COM-PHYS-2010-048, 2010.

[2] R. Nisius, "Interview," August 2011.

[3] T. Barillari, "Top mass measurements with the 'Jets ΔR Method'." Presentation, May 2010.

[4] M. Thomson, "Measurement of the mass of the W boson at LEP," *Topical Volume of Eur. Phys. J. C Direct*, 2004.

[5] R. Terrer, *Minimization with ROOT using TMinuit.* Stanford University, October 2001.