

An optimization of jet reconstruction for the Matrix Element Method in semileptonic $t\bar{t}H(b\bar{b})$ search using the ATLAS detector

Jeffrey Krupa, TRIUMF/CERN

August 20, 2016

Contact Information

jeffkrupa@gmail.com

Supervisor

Dr. Oliver Stelzer-Chilton, TRIUMF

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Oliver Stelzer-Chilton, whose direction, guidance, and insight was essential for the success of this project. Also invaluable was Alexander Held, a highly knowledgeable graduate student who fielded many questions, and provided me with continual support. I would like to also thank Dr. Jelena Jovicevic who generously provided me with her time and vast knowledge. I would also like to recognize the Institute of Particle Physics (IPP) and TRIUMF, who provided me with the unforgettable, inspiring, and humbling opportunity to participate in the CERN Summer School in Geneva, Switzerland.

Abstract

Run 2 of the LHC is probing physics at the TeV scale through pp collisions with center-of-mass energy $\sqrt{s} = 13$ TeV at unprecedented luminosities. One particular process that this promises to shed light on is the production of the Higgs boson in association with top quarks, $t\bar{t}H$. This report studies the reconstruction of jets from W boson decay in the $t\bar{t}H$ process. An algorithm to improve the reconstruction efficiency of the jets from W boson decay by recognizing their kinematic properties and applying specific requirements is introduced, implemented, and its effects are studied. The altered algorithm is shown to improve the separation between signal and background of the MEM calculations between 15% – 25%. The algorithm offers to potentially decrease the uncertainties in the calculated signal strength of the $t\bar{t}H$ process.

1 Introduction

1.1 Particle Physics

Particle physics is concerned with describing the most fundamental constituents of the universe as well as the state of the universe moments after the Big Bang. Particle physics seeks to answer fundamental questions in physics, such as the asymmetry of matter and antimatter in the universe, the unification of the forces of nature, and the Hierarchy Problem.

In pursuit of this, experimental particle physics employs particle accelerators to collide particles travelling near the speed of light. The energy from the collisions of these particles is then transformed into mass of new particles. These particles are detected by recording their decay products through interactions with materials in a particle detector. Using the sophisticated detector hardware and many layers of intricately designed software, the collision event can then be reconstructed.

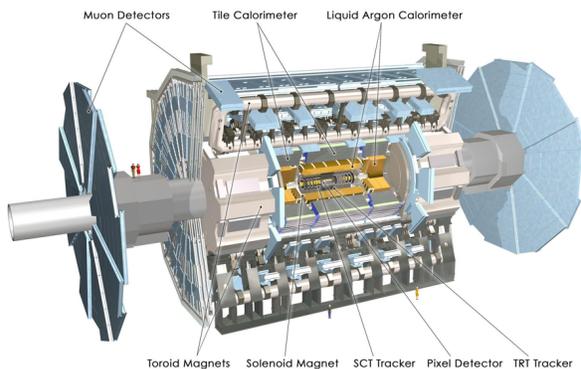


Figure 1: The ATLAS detector [1].

The world's largest and highest-energy collider is the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland. The collisions generated by the LHC allow for the probing of new physics at unprecedented energy scales, simulating the conditions of the early universe. Most recently, in run 2, the LHC has delivered protons that collide with a center-of-mass energy of $\sqrt{s} = 13$ TeV. One of the two multipurpose detectors designed to record these events at the LHC is the Toroidal LHC Apparatus (ATLAS) detector.

1.2 The ATLAS Experiment

The ATLAS detector is a 44 m long and 25 m diameter detector buried 100 m underground. A cross-section of the ATLAS detector is shown in Fig. 1. The innermost components of the ATLAS detector comprise the tracking system that detects the paths and momenta of charged particles curved by the magnetic field. Exterior to the tracker lie the calorimeters. The electromagnetic calorimeter is designed to detect the energy deposited by electrons and photons. The hadronic calorimeter is designed to detect the energy deposited by jets, which are cones of hadronizing quarks. Finally, the outermost region of ATLAS is the muon detector, which detects muons via the ionization of gas.

1.3 The Standard Model

The Standard Model (SM) of particle physics is the collection of all interactions of the known fundamental particles of physics. The particles encompassed within the SM are shown in Fig. 2a. The Higgs boson is a particle of particular interest to this study. Discovered in 2012 by the ATLAS [2] and Compact Muon Solenoid (CMS) [3] experiments, the Higgs boson has profound implications on the nature of mass of the elementary particles. In addition,

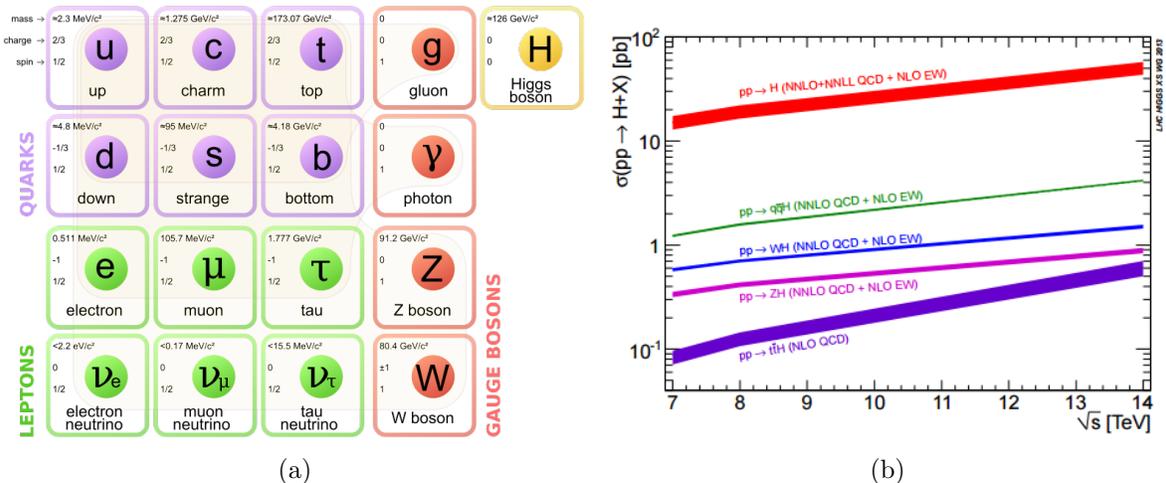


Figure 2: The (a) Standard Model of particle physics [4] and (b) cross sections of Higgs production methods [5].

it sheds insight into the early universe through electroweak symmetry breaking. Precision measurements of its properties are of great interest to particle physicists.

1.4 The Higgs boson and $t\bar{t}H$ Production

The cross sections of the known methods of producing the Higgs boson in pp collisions are shown in Fig. 2b.

The signal process studied here involves the search for the production of the Higgs boson in association with top quarks, denoted $t\bar{t}H$ [6], which has not yet been observed at the LHC. This is the lowest cross section Higgs production, seen in Fig. 2b. The signal $t\bar{t}H$ is interesting because it allows measurement and study of the Yukawa coupling between the top quark and Higgs boson, a measurement which has the potential to shed light on new physics. The decay channel $b\bar{b}$ is studied because it is the highest branching ratio of the Higgs boson. Ultimately, the signal events contain four jets originating from bottom quark hadronization, two jets originating from light quark hadronization, one detected lepton (electron or muon), and missing transverse momentum from an undetected neutrino.

A representative lowest-order Feynman diagram corresponding to the signal process is shown in Fig. 3. The main relevant background process in this study is $t\bar{t}$, which results in nearly the same decay products as the signal (specifically, it is irreducible with $t\bar{t} + b\bar{b}$). In addition, the background shows similar kinematics as the signal. Because of these facts, distinguishing between the signal and background based on their kinematical differences is not likely to provide adequate separation between the signal and background. It is clear, then, that a strong discriminating variable is required to separate the signal from background. A technique offering such discrimination is the matrix element method (MEM).

1.5 Relevant Requirements

Requirements, or cuts, are operations carried out on data and Monte Carlo (MC) simulations to isolate the events of interest. Their use is essential in collider physics in order to remove

the events caused by uninteresting scattering of the incident partons and pileup. Pileup are pp collisions in addition to the collision of interest [8]. These requirements are in addition to decisions at the trigger-level, wherein hardware and software decide whether or not to store the information from an event. Relevant variables on which cuts are performed in this study include the invariant mass of the jet pair M , defined in the relativistic limit in Eq. 1,

$$M = \sqrt{2 \cdot p_{T,1} \cdot p_{T,2} \cdot (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))} \quad (1)$$

pseudorapidity $\eta = -\ln[\tan[(\theta/2)]]$, polar angle ϕ , angular separation $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$, transverse momentum p_T and transverse energy E_T . Here, the quantities labelled with 1 and 2 refer to the quantities associated with jets 1 and 2, respectively. It is important to note that transverse quantities refer to the component of the quantity in the transverse plane, or orthogonal to the beam axis.

2 Matrix Element Method

2.1 Introduction

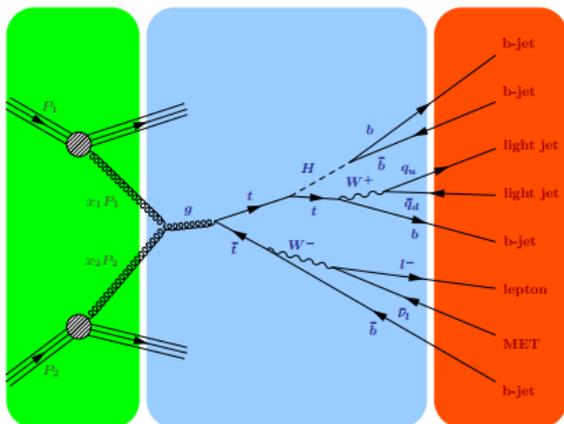


Figure 3: A representative lowest-order semileptonic ttH Feynman diagram [7]. The green column shows the incoming protons, the blue column depicts the decay chain, and the orange column displays the objects that the decay products are measured as.

The underlying principle of the matrix element method (MEM) is to calculate the likelihood that an event originated from a signal or a background process. The MEM calculation produces an MEM variable MEM_D1 , defined in Eq. 2,

$$MEM_D1 = \log \frac{L_{\text{sig}}}{L_{\text{bg}}} \quad (2)$$

where L_{sig} and L_{bg} are the likelihoods that an event is signal or background in origin, respectively. The separation power S between signal and background events provided by this variable is defined in Eq. 3,

$$S = \frac{1}{2} \sum \frac{(N_{\text{sig}} - N_{\text{bg}})^2}{N_{\text{sig}} + N_{\text{bg}}} \quad (3)$$

where N_{sig} and N_{bg} are the number of events classified as signal and background, respectively. For the MEM analysis, it is important to properly select the jets originating from each particle.

2.2 Truth- and Reconstruction-level Information

A valuable bit-wise integer available in our framework for simulated events is "jet truth-matching" information. This is a number assigned to a jet which is matched to the parton

it originates from if the two objects are within $\Delta R < 0.3$. Another important feature at the reconstruction level for the $t\bar{t}H$ analysis is the identification of jets originating from b quarks. Such a jet is denoted as being b-tagged. The algorithm used to quantify this is the mv2 (mv2c20 or mv2c10) algorithm, which assign each jet in a event a number between -1 and 1. If the jet is assigned a value greater than a specified threshold, the jet is deemed to be b-tagged.

2.3 Nominal Reconstruction Algorithm

The nominal selection algorithm is shown pictorially with 5 steps in Fig. 4a. For this analysis, it is first necessary to sort the events into a region where each contains no less than six jets, at least four of which originate from b quarks as prescribed by the mv2 algorithm. It is worth noting that the mv2 algorithms and tag rate function (TRF, another b-tagging algorithm instead providing boolean values) are in good agreement. This sorting is shown in step 1. A particular event is then examined. The lepton from this event is selected and removed, shown in step 2. The four jets with highest mv2 values are removed and associated to b quarks, shown in step 3. The invariant mass of each possible pair among the remaining jets, $M_{jj'}$ is calculated, and the pair jj' that minimizes $|M_{jj'} - M_W|$ is chosen as the jet pair originating from W decay, where $M_W = 80.4$ GeV is the mass of the W boson [6]. This is shown in step 4. The reconstructed event is shown in step 5.

One may argue that the pair that minimizes $|M_{jj'} - M_W|$ should be matched to the jets from W decay *prior* to selecting the 4 b-tagged jets; however this would result in a large number of calculations and opportunities for the algorithm to incorrectly select the pair from W. As a result, it is necessary to first reconstruct the b-tagged jets followed by the jets from the W decay.

For the studies here, which use the truth-matching information, the jet pair jj' selected using the nominal algorithm is correctly identified as originating from the W if one of the jets $\{j, j'\}$ is truth-matched to the leading (higher p_T) light quark, and the other is truth-matched to the subleading (lower p_T) light quark. The reconstruction efficiency, η_R , is defined in Eq. 4 using all relevant MC weights (which describe the significance that an event carries compared to the other events),

$$\eta_R = \frac{\text{events where jets truth-matched to jets from W decay are selected}}{\text{events in signal region}}. \quad (4)$$

2.4 Difficulties

In the ATLAS reconstruction software release Athena 20.1, the reconstruction efficiency using the nominal algorithm described previously was $\approx 27\%$. This is one of the most significant issues facing the MEM $t\bar{t}H$ analysis at the reconstruction-level, and it is the focus of this report. The poor reconstruction of jets originating from W decay may be caused by a significant portion of jets being lost due to the nominal requirements of $p_T > 25$ GeV and $|\eta| < 2.5$. Of course, recovering these lost jets is of great interest. The most appealing and simple solution would be to alter the p_T or $|\eta|$ requirements denoted $p_{T,cut}$ and η_{cut} , respectively, while limiting unwanted pileup and improperly tagged jets.

To this end, the requirements within the $t\bar{t}H$ source code were altered, and η_R was determined at various $p_{T,cut}$ and η_{cut} settings. The reconstruction efficiency as a function of η_{cut}

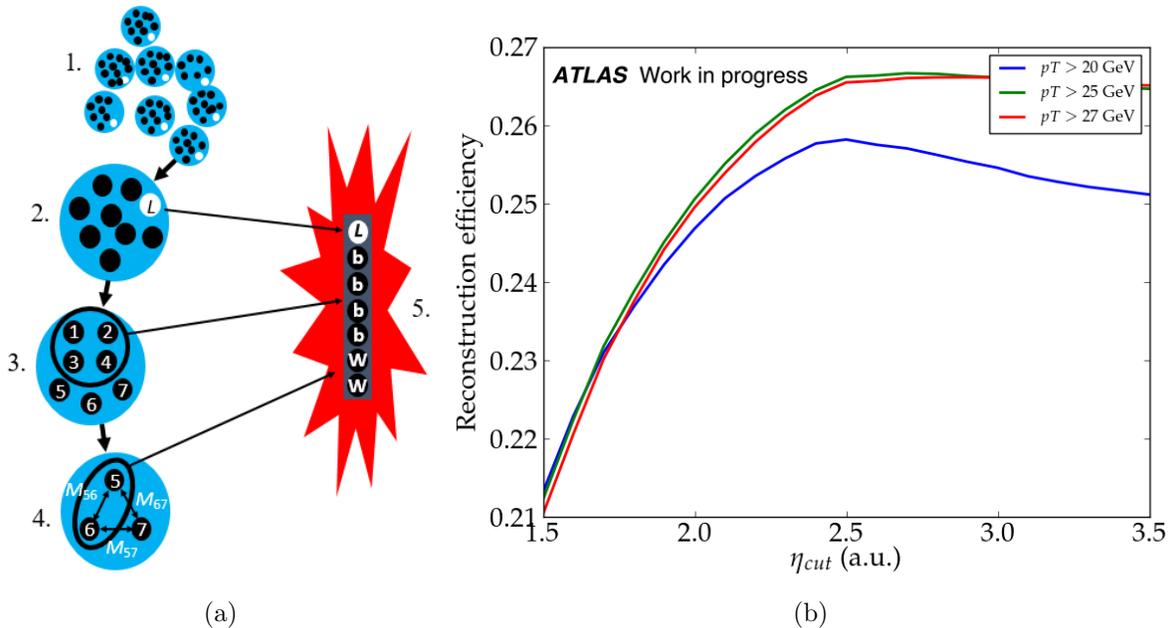


Figure 4: The (a) nominal reconstruction algorithm and (b) η_R as a function of η_{cut} for $p_T > 20$ GeV (blue), $p_T > 25$ GeV (green) and $p_T > 27$ GeV (red) requirements.

is shown in Fig. 4b for various p_T requirements (shown as curves with different colors). It is apparent that the nominal requirements are optimal in comparison to the loosened requirements. While studies show that loosening the requirements increases the absolute number of jets truth-matched to the jets from W decay, the contribution of pileup and contaminated events reduces η_R . In addition, looser requirements allow for larger combinatorics, and therefore a higher likelihood that the pair of jets from W decay is chosen incorrectly. Clearly, a different approach to improving η_R is required.

An added complication is the computational demand of the MEM calculations. A complete run over all systematics requires as much as two weeks of computing time. Fortunately, abbreviated calculations over nominal files taking only 6 hours may be performed to see the effects of changing the selection settings. These results were necessary to guide this optimization study, and will be presented in section 5.

3 Event Kinematical Profile

It is important and potentially useful to recognize the kinematic properties of the correctly identified jets from W decay (from truth-matching information), and compare them to the kinematic properties of the jet pairs selected by the selection algorithm. In other words, it is possible to compare the properties of the jets that *actually* originate from the W to the jets that are *chosen* to originate from the W. The properties of the truth-matched jets can then provide a reference for making the best possible choice when selecting the jets from W decay for situations where the truth-matching information is not available. The distributions showing these effects for a variety of kinematic variables are seen in Fig. 5.

From Fig. 5, certain differences are observed between the truth-matched pairs to W and

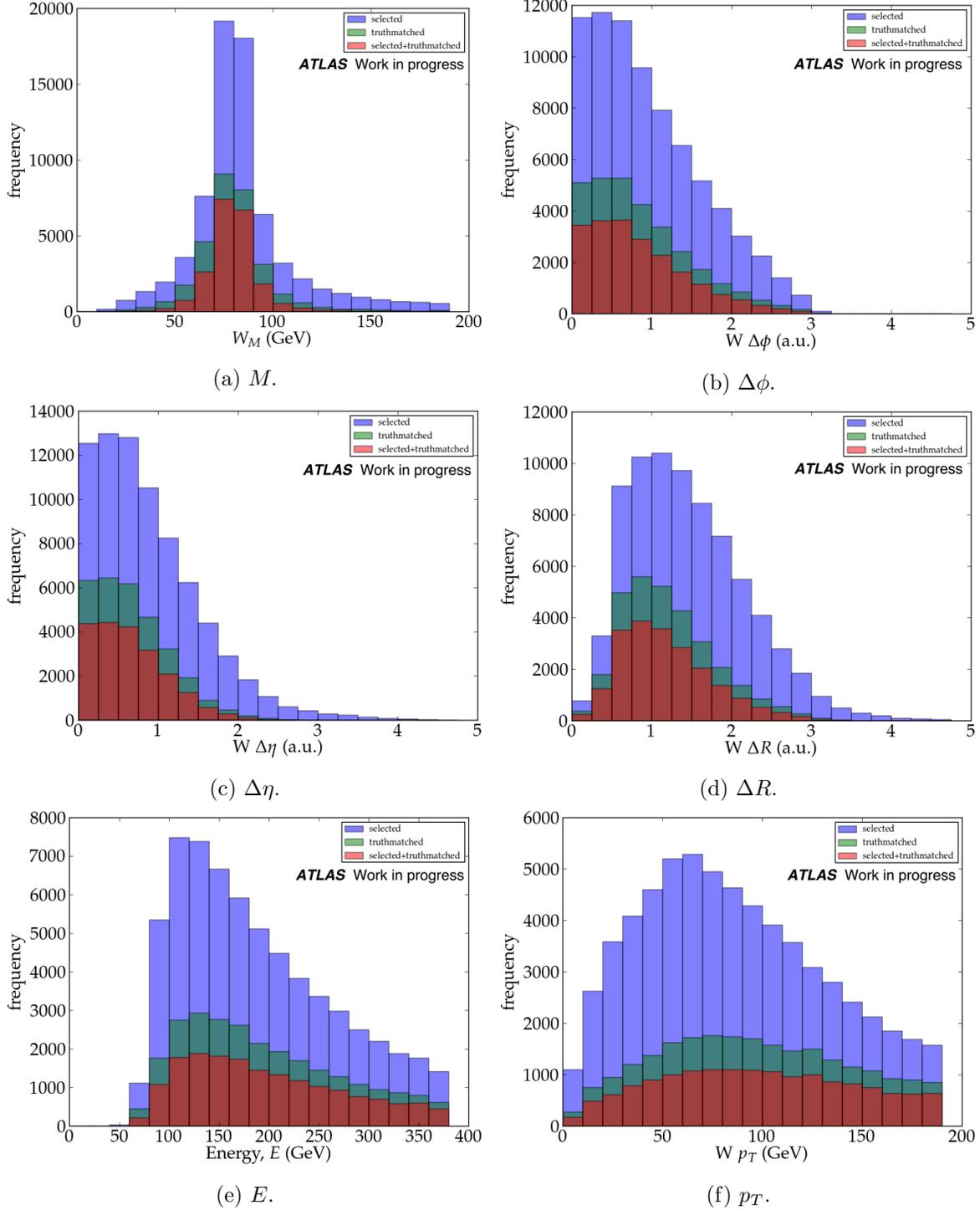


Figure 5: Various kinematic properties showing the jets from W selected using nominal algorithm (blue), truth-matched jets from W decay (green), and correctly identified jets from W decay (red).

the jet pairs selected by the algorithm. Notably, the truth-matched distribution for M is more sharply peaked near the W's mass than the selected distribution. Similarly, the truth-matched distribution for $\Delta\eta$ has a smaller tail than the selected distribution. As such, it is possible to perform cuts on these variables to preferentially select certain pairs that do originate from the W boson.

4 Jet Reconstruction Optimization

4.1 Introduction and Rationale

Recall that this study makes use of the truth-matching to the two quarks from W decay. Since a small fraction of events satisfy this truth requirement, the assumption that each event contains the jets from W decay is likely not justified for a majority of the events. One alternative is to discriminate between the events that are likely to contain the jets from W and the events that are unlikely to contain the jets from W decay for the MEM calculations. Each event then may be marked according to this event categorization.

The MEM calculations on the nominal file can then be performed by creating separation plots for both the regions of high- and low-purity, and a separation power can be obtained for both regions. A higher MEM separation power is expected in the high-purity region, which can help the fit in the $t\bar{t}H$ analysis by resulting in a more accurate signal strength measurement. This may ultimately provide a greater ability to claim whether or not the $t\bar{t}H$ process agrees with the SM.

4.2 Event Categorization

From Fig. 5a, it is apparent that the truth-matched jets from W decay follow a slightly different distribution than the jets selected and assumed to originate from the W decay according to the nominal selection algorithm. Outside of a certain range, it is highly probable that the jet pair chosen could not have originated from W. The altered algorithm introduced here involves placing requirements on M through two iterations of the nominal algorithm. Prior to the event being fully reconstructed, the b-tagged jets are removed, and the invariant mass of each remaining jet pair combination is calculated. If the jet pair jj' is the pair closest in mass to the W boson, and it satisfies the inequality $|M_{jj'} - M_W| \leq M_{cut}$, where M_{cut} is some arbitrary requirement on the invariant mass, the event is flagged as containing the jets from W decay and placed in a high-purity region. On the other hand, if the jet pair jj' does not satisfy this inequality, the event is flagged as not containing the W decay and placed in a low-purity region.

4.3 Selection Requirements

This method is faced with two constraints that must be satisfied. Firstly, it is required that each the high-purity and low-purity regions are both adequately large since a small region size will have no effect on the overall fit. Secondly, it is preferable for the high-purity region to contain many reconstructed jets from W decay, and the low-purity region to not contain many reconstructed jets from W decay. Unfortunately, with the small amount of events that contain truth-matched jets from W decay, in addition to the difficulties surrounding the

invariant mass requirement, it is likely not possible to achieve an entirely pure high-purity region and untainted low-purity region. It is possible however to vary the invariant mass for approximately equal event sharing between the regions. Indeed, it is likely worthwhile to distinguish between regions that are more or less likely to contain the jets from W.

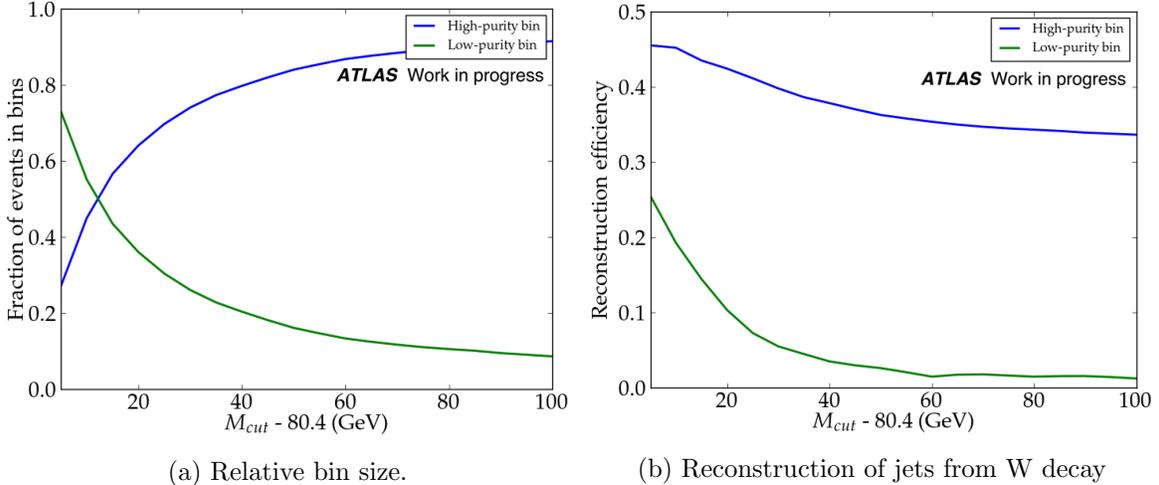


Figure 6: Plots showing the quantities as-labelled for high-purity region (blue) and low-purity region (green), as a function of invariant mass requirement.

4.3.1 Symmetric Event Categorization

A symmetric primary selection categorization places events in high- or low-purity regions using an invariant mass requirement that is symmetric on either side of M_W , as in $|M_{jj'} - M_W| \leq M_{cut}$, and is best suited for a symmetric distribution. The relative sizes of the high- and low-purity regions are shown with green and blue curves, respectively, in Fig. 6a. As expected, a restricted (i.e., small) M_{cut} results in equal sharing of events between the high- and low-purity regions, while a loosened (i.e., large) M_{cut} places all events in the high-purity region. The reconstruction efficiencies of jets from W decay both for the high- and low-purity regions are shown in Fig. 6b. Again as expected, a restricted M_{cut} results in higher η_R in both the high- and low-purity regions, respectively, while a loosened M_{cut} results in the nominal $\eta_R \approx 30\%$ in the high-purity region.

To test the impact on the separation provided by the symmetric M_{cut} , separation plots were generated at $M_{cuts} = 5$ GeV, 11 GeV, and 80 GeV, in order to gain insight into the impact of the purity and statistical requirements on the MEM separation power. The results will be described in section 5.

4.3.2 Asymmetric Event Categorization

An asymmetric primary selection categorization places events in high- or low-purity regions using an invariant mass requirement that is asymmetric on either side of M_W . This approach may be justified since the distribution of the truth-matched jets from W decay shows different behaviour on either side of M_W , as seen in Fig. 5a. The density plots in Fig. 7a and 7b show η_R in the high- and low-purity regions, respectively, as a function of the requirement

above M_W , $M_{uppercut}$, and the requirement below M_W , $M_{lowercut}$. The relative size of the high-purity region is shown in Fig. 7c.

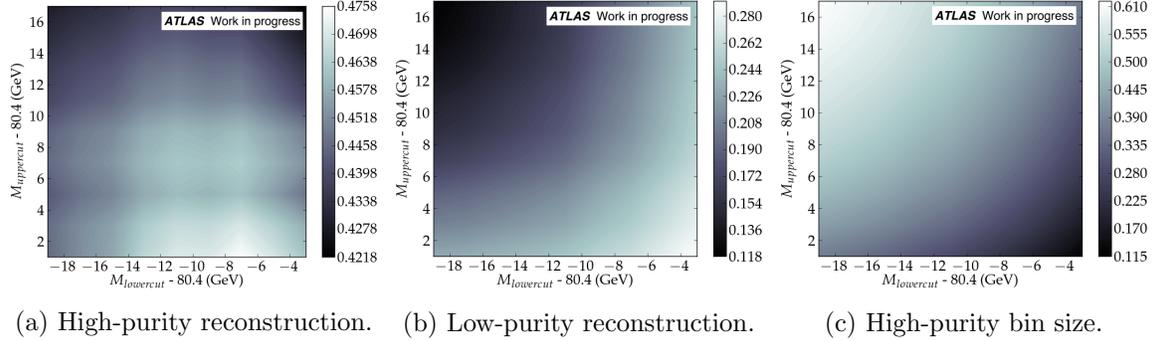


Figure 7: Density plots showing the quantities as-labelled, as a function of $M_{uppercut}$ and $M_{lowercut}$.

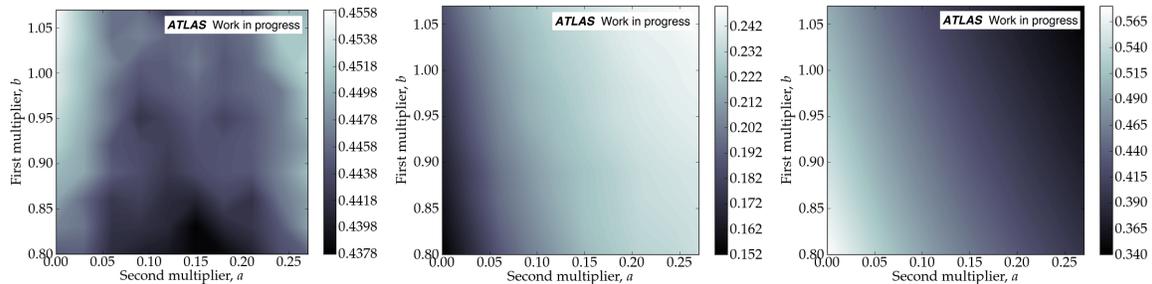
The statistical requirement for the fit dictates that the high- and low-purity regions must approximately equally share the events. As a result, the asymmetric categorization is limited to a specified region in Fig. 7a. The increase in η_R using this approach is marginal. Notably, an increase in η_R of 0.5% is recorded when $M_{uppercut} = 8$ GeV and $M_{lowercut} = -12$ GeV, while maintaining equal sharing of the events. In addition, Fig. 7c shows contour cross-sections akin to a quarter-circle, unfortunately not allowing for regions of particularly diminished reconstruction of jets from W. Ultimately, the required regime laid out by the statistical constraint of event-sharing shown in Fig. 7c does not allow for any significant increase in η_R in the high-purity region or decrease in reconstruction in the low-purity region. This result shows that, despite the asymmetric M distribution, an asymmetric M_{cut} is not worthwhile. As such, it is not of further use to this study.

4.3.3 The Other Jet Pairs

It is possible that judging an event by the jet pair closest in invariant mass to the W is not the optimal approach. Insofar, I have claimed that if an event does not contain a jet pair sufficiently close to the W (i.e., within $M_W \pm M_{cut}$), it can be assumed to not contain the jets from W decay. If the lack of the jets from W is indeed an aspect of certain events, and it is possible to choose jet pairs that are not closest to M_W , it may be beneficial to consider numerous jet pairs when deciding on the categorization of an event. To simplify the analysis and place a constraint on the number of free parameters, only the two jet pairs closest in invariant mass to the W are considered.

Events are categorized into high- and low-purity regions depending on whether they satisfy the inequality $|b \times M_{jj'} + a \times M_{jj''} - M_W| \leq M_{cut}$ where jj' is the pair closest-to-W, jj'' is the pair next-closest to-W, and b and a are multiplicative constants. However, since it is unclear which multiplicative factors are necessary for optimal reconstruction, it was necessary to calculate η_R for a variety of combinations of a and b parameters. The resulting plots in the same format as Fig. 7 are shown in Fig. 8.

It is clear that this approach to considering the event as a whole does not yield sustan- tively larger η_R , as the optimal setting is the nominal setting, located at $b \approx 1$ and $a \approx 0$. Of course, increasing b at fixed $a = 0$ yields larger η_R , but this effect is simply a more restricted



(a) High-purity reconstruction. (b) Low-purity reconstruction. (c) High-purity bin size.

Figure 8: Density plots showing the quantities as-labelled, as a function of multiplicative constants a and b .

M_{cut} , as previously described. There are certain regions, for example $b \approx 0.9$ and $a \approx 0.08$, which show comparable η_R to the symmetric M_{cut} setting, however they are similar in η_R to the nominal setting. Perhaps it is necessary to include more jet combinations in the ranking of jet pairs instead of the first two, however that was not attempted here. In any case, this approach is not of further use to this study.

4.4 Jet Selection

The next component to the analysis here involves determining a way of choosing the jet pair from W. According to the distributions in Fig. 5, it may be possible to select jet pairs corresponding to certain kinematical properties. To this end, the jet pairs truth-matched as originating from W decay, ranked by increasing absolute mass away from M_W , are shown in Fig. 9. Without additional selections in place, only the entries in the first bin (the closest-to-W) are selected. Selecting the truth-matched jets in the second or higher bins in Fig. 9 is worthwhile. It is worth noting that a boosted decision tree (BDT) could be very useful here.

It is clear from the first bin in Fig. 9 that the jet pair closest-to-W is very likely to be the jet pair that is also truth-matched to originating from the W (shown by the first bin being the largest). However, it is also apparent that occasionally, in about 33% of jet pair selections, the jet pair truth-matched to the W is not the nominally-chosen closest-to-W (shown by subsequent bins being non-zero). It is of interest, then, to attempt to recapture these jets which usually remain unchosen by the application of specific requirements.

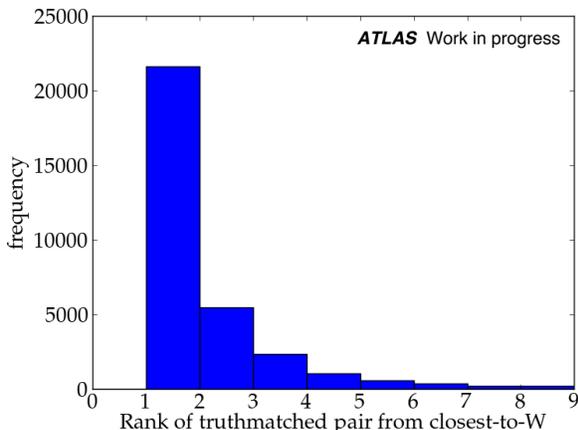
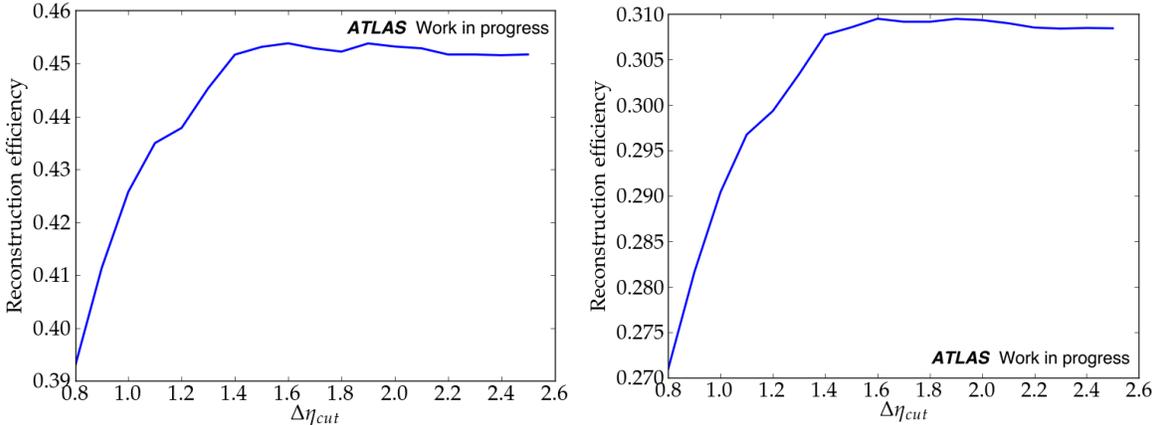


Figure 9: Truth-matched jet pairs originating from W in order of distance from M_W .

4.4.1 $\Delta\eta$ Selection

Firstly, it is not clear whether preferentially selecting jets exhibiting a distribution with a peak and a tail on only one side (e.g., $\Delta\eta$ in Fig. 5c), or jets exhibiting a distribution with a peak with tails on both sides (e.g., p_T in Fig. 5f) will be more effective in increasing η_R of jets from W decay. For instance, the distribution of $\Delta\eta$ shows that jet pairs chosen with $\Delta\eta > 2.0$ are chosen improperly, and as such, jet pairs can preferentially be chosen to satisfy $\Delta\eta < 2.0$. However, it is apparent that the fraction of selected jets beyond this region is a small fraction of the total jets selected, so a marginal improvement is expected. It is important to note that, while the $\Delta\eta$ distribution shows differences between the truth-matched and selected distributions, the $\Delta\phi$ distribution does not. As a result, the often-used ΔR distribution was not as useful, since the effect of the $\Delta\eta$ distribution is washed out by the addition of the $\Delta\phi$ distribution in quadrature.

An algorithm to investigate this may be implemented by arranging the jet pairs in increasing M separation from M_W . If the jet pair in question does not satisfy a particular requirement $\Delta\eta_{cut}$, the next-closest-to-W pair is checked. If this jet pair does satisfy the $\Delta\eta_{cut}$, the algorithm stops and this pair is chosen. If the jet pair does not satisfy the $\Delta\eta_{cut}$, the next pair is checked, and the algorithm continues. The code for this algorithm is implemented, and the results are displayed in Fig. 10, which shows η_R as a function of $\Delta\eta_{cut}$ in the high-purity region for two M_{cut} settings.



(a) $M_{cut} = 11$ GeV.

(b) $M_{cut} \rightarrow \infty$.

Figure 10: Plots showing the reconstruction efficiency in the high-purity region as a function of $\Delta\eta_{cut}$ for settings as-labelled.

From Fig. 10 a small increase in η_R of about 0.5% is gained with choice of $\Delta\eta_{cut} \approx 1.5$. It is worthwhile to note that, due to an apparent increase in sensitivity to the jets from W decay at the truth-level arising in Athena 20.7 files, the earlier-seen 1.5% increase in reconstruction efficiency was later severely diminished. In any case, the effectiveness of this requirement was tested by calculating MEM likelihoods to determine its impact on separation power. These results will be described in the results section. Finally, it is worth noting that similar plots for ΔR were generated; however jet selection in this way with ΔR had no effect.

4.4.2 p_T Selection

An example of a peaked distribution with tails on both sides of the peak is p_T , shown in Fig. 5f. In addition, this distribution seems appealing for optimizing the selection of jets from W decay, as the truth-matched jets from W exhibit a peak separated by approximately 10 GeV from the selected jets from W decay. This shift is appealing because it displays the nominal algorithm's bias in selecting jets with insufficient p_T . However, this separation is a small fraction of the overall span of the distribution. The same algorithm as above for $\Delta\eta$ is implemented, wherein the jet pair selected by the algorithm was rejected if it does not satisfy a $p_{T,cut}$. The results are displayed in Fig. 11, which shows η_R as a function of $p_{T,cut}$ in the high-purity region for two M_{cut} settings.

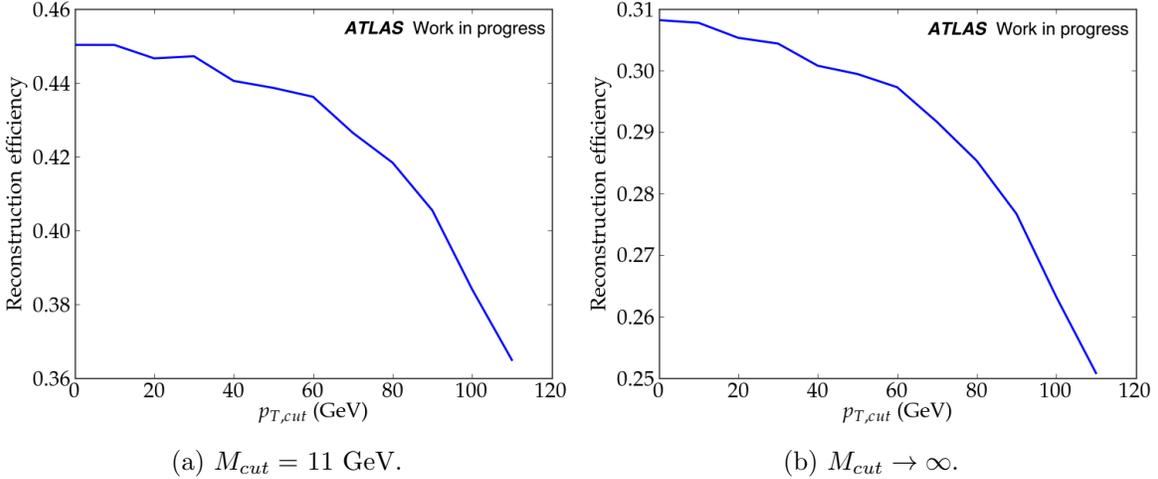


Figure 11: Plots showing the reconstruction efficiency in the high-purity region as a function of $p_{T,cut}$ for settings as-labelled.

From Fig. 11, it is apparent that any p_T requirement diminishes η_R . Therefore, this selection technique is not of further use to this study.

4.5 Selecting Jets With ΔR

It is possible that selecting jets primarily by their invariant mass is not the method yielding optimal reconstruction efficiencies. A possible alternative to primarily selecting the jets is yielded by the distribution of ΔR , shown in Fig. 5d. In this algorithm, the jets pairs are ordered in increasing angular separation ΔR from the peak of the truth-matched distribution, at $\Delta R = 0.8$ (i.e., minimizing $|\Delta R_{jj'} - 0.8|$). Then, in this order, it is determined whether the jet pair jj' has a mass within a window M_{cut} from M_W (i.e., whether it satisfies $|M_{jj'} - M_W| \leq M_{cut}$). If the jet pair does satisfy this requirement, the algorithm stops and this pair is chosen. If the jet pair does not satisfy this requirement, the next pair is checked, and the algorithm continues. The code for this algorithm is implemented, and the results are displayed in Fig. 12, which shows η_R as a function of M_{cut} .

It is apparent from Fig. 12 that $\eta_R \approx 29.5\%$ is the largest η_R yielded by this selection algorithm. This occurs at $M_{cut} \approx 25$ GeV. However, this is slightly smaller than η_R obtained using the previously described invariant mass algorithm. Therefore, the ΔR selection method is not of further use to this study.

5 Results

It is important to note that at the time of the completion of this report, a full fit showing the results of this analysis was not available because the MEM evaluation was still running.

5.1 Effect of Event Categorization

Separation plots for the nominal file run over setting $M_{cut} = 11$ GeV are shown in Fig. 13. The separation plot without splitting regions is shown in Fig. 13a, while the separation plots only containing the events in the high- and low-purity regions are shown in Fig. 13b and 13c respectively. It is important to recognize that these results do not show the signal strength, as a full MEM fit is required. However, they do provide a separation power. Recall that a larger separation power is expected to yield smaller uncertainties in the signal strength, which is beneficial for precision physics measurements, and therefore comparisons between experimental and expected values.

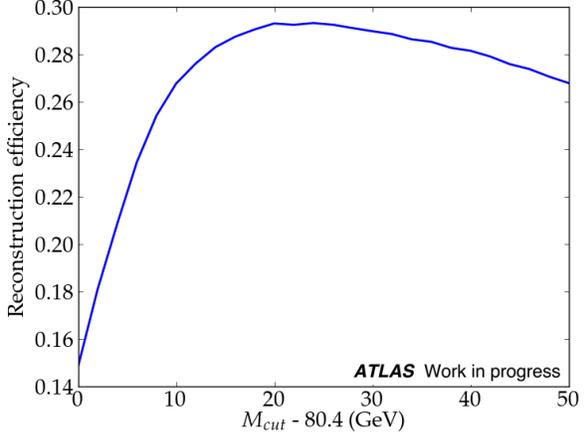
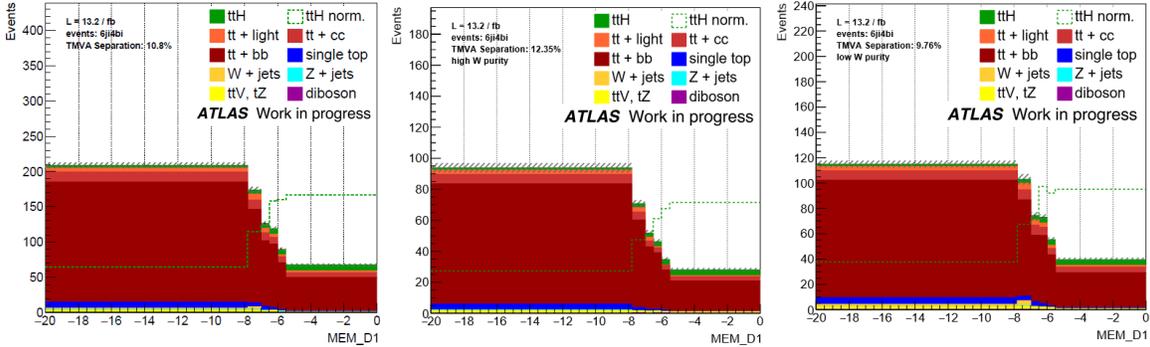


Figure 12: The reconstruction efficiency when the jet pairs are ranked sequentially by ΔR from the W, and rejected if a closeness-to-W mass requirement M_{cut} is not satisfied.



(a) Inclusive bin: $S = 10.8\%$. (b) High-purity bin: $S = 12.4\%$. (c) Low-purity bin: $S = 9.8\%$.

Figure 13: MEM separation plots for $M_{cut} = 11$ GeV for regions as-labelled.

A relative improvement in separation, calculated with $\frac{S_{\text{High-W bin}} - S_{\text{Inclusive bin}}}{S_{\text{Inclusive bin}}}$, of 15% is gained by splitting the events into high- and low-purity regions. This gain is attributed to larger η_R in this region. On the other hand, a decrease in separation in the low-purity region is evident. However, this is expected, since the low-purity region exhibits a low η_R , and therefore a diminished ability for the MEM calculations to discriminate signal from background. It is

M_{cut}	No bins	High-W purity	Low-W purity	Relative Improvement
$M_W \pm 5$ GeV	9.67%	13.77%	9.36%	42%
$M_W \pm 11$ GeV	10.44%	13.21%	9.75%	26%
$M_W \pm 80$ GeV	10.17%	10.89%	9.28%	7%

Table 1: The first three columns show the separation results in both the high- and low-purity regions for various event categorization settings. The right-most bin shows the relative increase in separation in the high-purity bin, calculated with $\frac{S_{\text{High-W bin}} - S_{\text{Inclusive bin}}}{S_{\text{Inclusive bin}}}$.

important to recognize that these two conflicting effects contribute to the overall MEM fit; as such, a full fit is required to see the combined effect, and will be performed. It is anticipated that the net effect will be an increase in the separation strength, implying a decrease in uncertainties of the output signal strength.

More results of the separation power using various splitting settings for the high- and low-purity regions are shown in Tab. 1. In each case, a significant relative increase in separation in the high-purity region is observed, balanced by a somewhat less significant decrease in separation in the low-purity region. The separation plots, as seen in Fig. 13, are not included, since the only quantity of interest in this study is the separation power.

It is apparent from Tab. 1 that there is significant fluctuation in the separation provided by the MEM calculations, due to the integration settings currently used. The best measure of this method's effectiveness, then, is the relative improvement column showing the difference for each setting, as opposed to direct comparisons between the separation provided by the settings. While the relative improvement is large for restricted (i.e., small) M_{cut} settings, such as $M_{cut} = 5$ GeV, it is possible that these regions do not satisfy the statistical requirement previously outlined. On the other hand, relative improvement almost vanishes for loosened (i.e., large) M_{cut} settings, such as $M_{cut} = 80$ GeV. This lends itself to a balanced M_{cut} requirement in the region of $10 \text{ GeV} \leq M_{cut} \leq 30 \text{ GeV}$, where events are reasonably shared between the high- and low-purity regions (as shown in Fig. 6a), but an improvement is still observed in the MEM separation between signal and background.

A setting of interest not shown here is $M_{cut} = 27$ GeV. Here, η_R in the low-purity region is limited to $\approx 5\%$. Any event categorization more restricted than this causes a large increase in η_R in the low-purity region, resulting from jets from W decay being improperly placed in the low-purity region. The benefit, however, is a sizeable 10% increase in η_R from the nominal setting of $M_{cut} = \infty$. Any cut more restricted than this causes a sub-linear increase in jets from W decay being properly placed in the high-purity region.

5.2 Effect of Jet Selection

The $\Delta\eta_{cut}$ was the only jet selection technique that was found to improve η_R in this study. When the MEM likelihoods were evaluated, a relative improvement in separation power of approximately 3% was obtained. This is a marginal effect, and not decidedly above the statistical fluctuations in the separation.

6 Discussion and Conclusion

This report outlined a study in optimizing the reconstruction of jets from W decay for use in a matrix element method analysis of the production of the Higgs boson in association with top quarks. The nominal reconstruction algorithm of the analysis, which reconstruct the lepton, bottom quarks, and jets from W decay, was introduced. Poor reconstruction of the jets from W decay was noted, and attributed to jets truth-matched to the W decay lost due to p_T and η requirements. However modifying these parameters was deemed to be ineffective due to increased pileup, decreased b-tagging efficiency, and increased combinatorics. A modification to this algorithm was introduced, wherein an event was placed in distinct regions of either high- or low-purity of jets from W decay depending on the properties of the jets present in the event.

This approach offered significant increases in η_R in the high-W purity region at the expense of improperly reconstructed jets in the low-W purity region. A variety of methods to split the events into bins were described in an attempt to limit this improper reconstruction, including asymmetric M_{cut} 's and the consideration of more jet combinations. In addition, jet selections were introduced and implemented by ordering the jets in increasing separation from the mass from the W decay, and rejecting pairs if they did not satisfy a requirement on the variable in question. Unfortunately, the latter requirements were not determined to be very useful. They generally exhibited decidedly small improvement in η_R , with exception of marginal improvement offered by a requirement placed on $\Delta\eta$.

Using this modified approach to the analysis, preliminary results including separation powers of were obtained by running MEM calculations over the $t\bar{t}H$ signal and all backgrounds. The results ranged widely in effectiveness. Statistical- and purity-related restrictions on M_{cut} in large part dictate the choice of settings. A benchmark setting of $M_{cut} = 11$ GeV was shown to increase separation power from 15% – 25%. A full fit must be performed to obtain a complete understanding of the relation between the increase in separation power in the high-purity bin, and the decrease in separation power in the low-purity region.

More studies could be pursued to achieved a stronger impact for the jet selection. In particular, adapting BDT algorithms to choose or reject jet pairs from W depending on particular kinematic properties would be highly useful. The sorting algorithm could also be refined further. A series of requirements could be used to calculate a "score" for an event, based on a series of kinematical requirements. The event could be placed in the high- or low-purity region according to whether it satisfies this score. This however is a non-trivial optimization and computing task with many free parameters to calculate. In addition, a different and more complicated sorting algorithm could be considered. For example, considering the jet multiplicity in addition to the invariant mass of the jet pairs may be of use. This is because an event with many jet combinations is unlikely to correctly identify the pair of jets from W decay.

In conclusion, this report showed that there is certainly potential to increase the separation power of MEM calculations by increasing the reconstruction efficiency of jets from W decay. This may be of use in decreasing calculated uncertainties in the signal strengths in the semileptonic $t\bar{t}H$ analysis at ATLAS, allowing for better precision physics, and by extension tests of the SM.

References

- [1] “Structure and Function of the ATLAS Detector.” <https://kjende.web.cern.ch/kjende/en/atlas.htm>. Accessed: 2016-08-05.
- [2] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, pp. 1–29, Sept. 2012.
- [3] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, pp. 30–61, Sept. 2012.
- [4] “Standard Model of Elementary Particles.” https://en.wikipedia.org/wiki/Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg. Accessed: 2016-08-05.
- [5] M. Carena *et al.*, “Status of Higgs Boson Physics,” vol. 716, p. 12, Nov. 2013.
- [6] G. Aad *et al.*, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into in collisions at with the ATLAS detector,” *European Physical Journal C*, vol. 75, p. 349, July 2015.
- [7] O. Nackenhorst *et al.*, *Search for the Standard Model Higgs boson produced in association with tt and decaying into bb at 8 TeV with the ATLAS detector using the Matrix Element Method*. PhD thesis, Gottingen U., May 2015. Presented 08 Jun 2015.
- [8] Z. Marshall and the Atlas Collaboration, “Simulation of pile-up in the atlas experiment,” *Journal of Physics: Conference Series*, vol. 513, no. 2, p. 022024, 2014.