CERN Summer Student Project 2018

Final Report

Reconstructing semileptonic top-pairs using different strategies

Konstantinos Christoforou

Supervisors
Teng Jian Khoo, Josh McFayden

September 2018

Abstract

Since at the LHC[1] are taking place $p\bar{p}$ collisions in the center of ATLAS[2] detector with energies of $\sqrt{s} = 13$ TeV, the production rate of top quarks is higher than ever. Due to its massive nature, top quark is the last discovered quark and nowadays, with the potential given by the LHC, many top production processes are accessible for precise and extended measurements of its properties. Besides of research interest as individual, better understanding of tops' nature is important in searches for new particles because SM processes involving top quark production, consist a very strong background for new physics. One of the most interesting top production processes, is the semileptonic $t\bar{t}$ channel, the study of which could be challenging. This project focuses on the comparison of 3 different top reconstruction techniques using jets given by 2 different jet reconstruction algorithms. This report will provide reader with information concering why a good MET guess is decisive, which reconstruction technique seems to be better for unfolding this process and if, at the end, the accuracy of $t\bar{t}$ reconstruction strongly depends on jets' reconstruction accuracy.

Contents

1 Introduction

The semileptonic $t\bar{t}$ channel, considering events with e or μ, represents $\sim$ 30 % of all possible $t\bar{t}$ decay channels. This project focuses on the case with four fully resolved jets in the final state, 2 light jets coming from the hadronically decaying W (W_{had}) and 2 b-tagged originating from the 2 tops' decay. The first challenging part of the event reconstruction is the reconstruction of the leptonically decaying W(W_{lep}), because of the neutrino existence in the final state. Since neutrino interacts very weakly with the matter, we cannot detect using the ATLAS detector system, which means that there is a missing information about the event. Taking this into account, the W_{lep} reconstruction has to define a way to resolve the ambiguity in the neutrino measurement. For that, some ways of doing so have been explored.

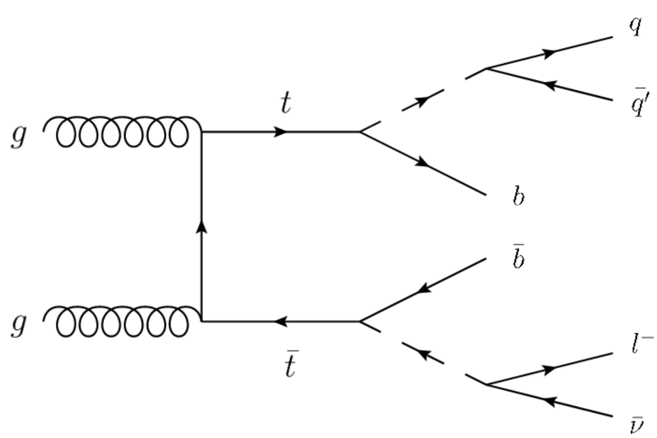


Figure 1.1: *Semileptonic $t\bar{t}$ process*

Thereafter reconstructing the W_{lep}, the second and maybe even more challenging part, is the reconstruction of W_{had} and matching it with the corresponding b-jet in order to combine them and finally reconstruct the t_{had}. Matching the W_{lep} with its corresponding b-jet, reconstructing the t_{lep}, has to be done at the same time. For the needs of this demanding procedure, 3 different top reconstruction techniques have been tested: the PseudoTop approach, χ^2 method and a likelihood-based method, KLFitter, in order to find out which one is the most efficient. At the end, as a last but not least part of this project, we are going to investigate if the accuracy of $t\bar{t}$ reconstruction is strongly depends on the jets' reconstruction accuracy. This can be achieved by comparing the efficiency of "well reconstructed events" by comparing 2 different jet reconstruction algorithms, the Topological clustering algorithm and the ParticleFlow algorithm.

2 Neutrino Reconstruction

One widely used way of doing the neutrino reconstruction and, by extension, the W_{lep} reconstruction is under assumptions because of the missing neutrino information.

First assumption is that neutrino's p_T should be equal to the missing transverse energy, MET. MET is set equal to the negative vector sum of p_T for all detected objects in a particle detector.

$$\vec{\not{E}}_T = -\sum_{i}^{N_{PF}} \vec{p}_{T,i}, \qquad MET = |\vec{\not{E}}_T| \tag{1}$$

According to the S.M., neutrinos are massless particles, so having the p_T and the mass of the emitted neutrino, the last piece of its 4-vector is the ${p_z}^{\nu}$. So, it's a straightforward special relativity problem to define the invariant W mass in terms of the knowns and the unknown ${p_z}^{\nu}$, for which is possible then to solve for. The calculations assume that the lepton mass is very small with respect to its momentum ($m_{\nu_l} = 0, m_l = 0 \Rightarrow {E_{\nu,l}}^2 = \vec{p}^2_{\nu,l}, E_{\nu,l} = |\vec{p}_{\nu,l}|$)

$$(p^W)^2 = (p^\nu + p^l)^2 \Rightarrow {m_W}^2 = (E_\nu + E_l)^2 - (\vec{p}_\nu + \vec{p}_l)^2$$

$$\Rightarrow \tfrac{{m_W}^2}{2} = |\vec{p}_\nu||\vec{p}_l| - \vec{p}_\nu \bullet \vec{p}_l$$

$$\Rightarrow \tfrac{{m_W}^2}{2} = \sqrt{({p_T}^l)^2 + ({p_z}^l)^2}\sqrt{({p_T}^\nu)^2 + ({p_z}^\nu)^2} - ({p_T}^l \bullet {p_T}^\nu + {p_z}^l {p_z}^\nu)$$

$$\Rightarrow ({p_T}^l)^2({p_z}^\nu)^2 - [{m_W}^2{p_z}^l + 2({p_T}^l \bullet {p_T}^\nu){p_z}^l]{p_z}^\nu$$
$$+ ({p_z}^l)^2({p_T}^\nu)^2 - (\frac{{m_W}^2}{2})^2 - {m_W}^2({p_T}^l \bullet {p_T}^\nu) + ({p_T}^l)^2({p_T}^\nu)^2 - ({p_T}^l \bullet {p_T}^\nu)^2 = 0$$

The above equation is a quadratic equation so the formula of solving it for the unknown ${p_z}^\nu$ is presented below.

- $a = ({p_T}^l)^2$
- $b = -[{m_W}^2{p_z}^l + 2({p_T}^l \bullet {p_T}^\nu){p_z}^l]$
- $c = ({p_z}^l)^2({p_T}^\nu)^2 - (\frac{{m_W}^2}{2})^2 - {m_W}^2({p_T}^l \bullet {p_T}^\nu) + ({p_T}^l)^2({p_T}^\nu)^2 - ({p_T}^l \bullet {p_T}^\nu)^2$
- $\Delta = b^2 - 4ac$

$$\boxed{{p_z}^\nu = \frac{-b \pm \sqrt{\Delta}}{2a}}$$

Of course, solving a quadratic equation, 2 solutions may come up. So, under convention, the one with the lowest absolute value was kept. Having the last missing piece of the neutrino, the reconstruction of W_{lep} could be easily calculated by combining the 4-vector of the reconstructed lepton and neutrino. Although, $\sim 34\%$ of events yield complex solutions for ${p_z}^{\nu}$, which poses a problem for subsequent top reconstruction.

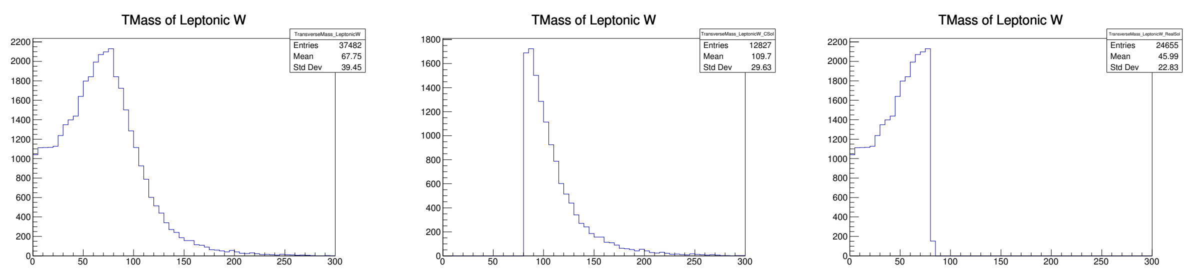
In order to investigate this, we had to check various observables, such us the transverse mass of W_lep, which provides information about its mass, without relying on the longitudinal component of the neutrino momentum. The transverse mass of a particle decaying into 2 different particles is defined as:

$$ {m_T}^2 = (E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2 \quad (2) $$

In our case, the lepton and the neutrino were assumed to be massless. As a result, the m_T^W could be defined by the p_T of the 2 particles and the θ angle between them, in the transverse plane (eq.3).

By splitting the cases to those that have and do not have an Imaginary part, as it seems from Fig.2.1, events where $m_T^W \leq m_W$ was violated, results ${p_z}^{\nu}$ to be complex number, as the breaking of this inequality results to nonphysical outcomes.

$$ (m_T^W)^2 = 2 * p_{T,l} * p_{T,\nu} * (1 - \cos\theta) \quad (3) $$



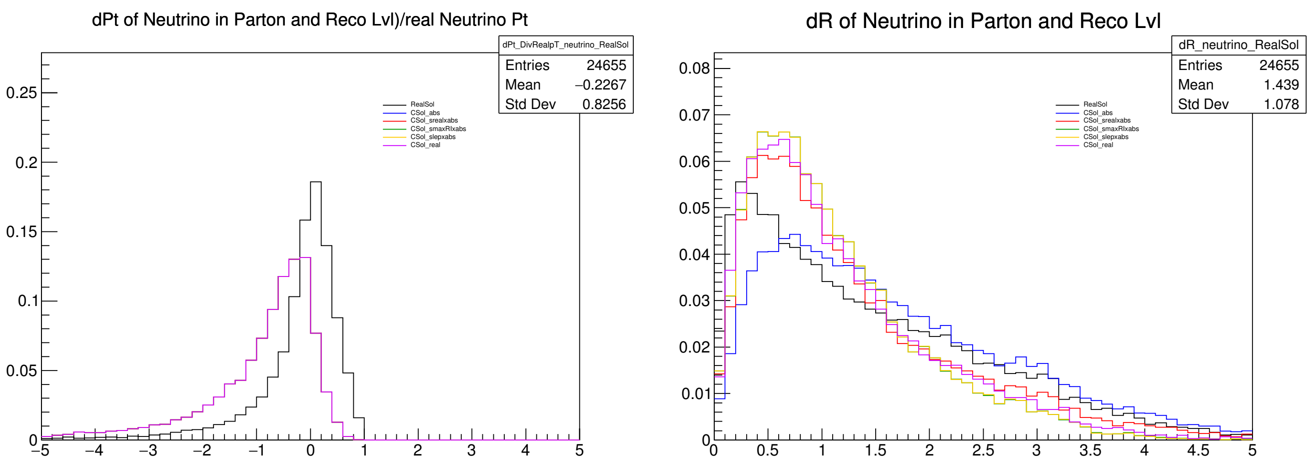
***(a)** Inclusive* ***(b)** Complex Solutions* ***(c)** Real Solutions*

Figure 2.1: $m_T^{W_{lep}}$

2.1 Neutrino Investigation

Constituting the 1/3 of the total number of events, throwing away the events where complex solutions come up for the ${p_z}^{\nu}$ is not optimal. The origin of these unphysical results should be found and of course what information could be recovered from the corresponding events. Thus, many different approaches were tested on how to deal with complex solutions such as using just the real part of the number (represented by magenta color) the magnitude of the complex number (represented by blue color), trying the same but multiplying the magnitude either by the real part's sign (represented by red) or by the sign of max(real,Imag) (green) or by the sign of the lepton's p_z (yellow). As it turns out, the last approach (sign[lepton.pz]*magnitude) seems to make the best neutrino guess, in terms of ΔR between the real neutrino and the reconstructed one (Fig.2.2b).

- Real Solutions
- Complex Solutions, $p_z{}^\nu = |(-b \pm \sqrt{\Delta})/2a|$
- CSol, (sign of real part)*magnitude
- CSol, sign[max(real,Imag)]*magnitude
- CSol, sign[lepton.pz]*magnitude
- CSol, $[p_z{}^\nu].Real$



(a) $(p_T^{real} - p_T^{reco})/p_T^{real}$

(b) ΔR*(real neutrino,reco)*

Figure 2.2

According to the Fig.2.2a, which represents the difference in p_T between the real and the reconstructed neutrino, someone can easily understand that the distribution for the complex solutions is shifted to the left, in respect of the real solutions' one. Meaning that the majority of the events where the MET was overestimated, lead to complex solutions for the $p_z{}^\nu$ and consequently to unphysical results. A closer look to Fig. 2.2b, comparison between the distribution for real solutions (black-line) and complex solutions (for example the sign[lepton.pz]*magnitude), indicates that although for real solutions is closer to 0, the tail is quite bigger (wider distribution). So, recovering the information of the complex solutions may be interesting and useful after all. For the moment, keeping in mind that the pursuit of a better MET guess is mandatory, for the needs of this project, all the studies which follows, correspond to events where the ${p_z}^\nu$ was a real number.

3 The 3 different reconstruction techniques

Three different methods were used for the top-pair reconstruction

1. PseudoTop approach
   PseudoTop approach combines the detected lepton and the MET to reconstruct the W_{lep}, as was presented on Section 2. Afterwards, is trying to choose the best light dijet combination in order to reconstruct the W_{had}. The best combination is the one that gives the closest m_{jj} to the well-known value of m_W. Having the 2 W's, this approach assumes that the closest b-jet in ΔR to the W_{lep} came from the same top (t_{lep}) and ,respectively, the closest one to the W_{had} should originate from the same disintegration, top_{had}'s decay.

2. chi2 minimization method [[3](Section 5)]

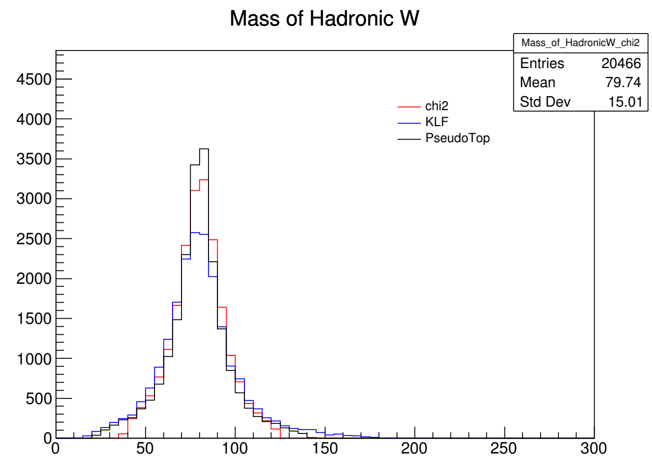
   - $\chi^2 = \frac{(m_{j_1 j_2} - m_W^{MC})^2}{\sigma^2_{m_W^{MC}}} + \frac{(m_{b_1 j_1 j_2} - m_{b_2 l \nu_l})^2}{\sigma^2_{\Delta m_{bjj}}}$

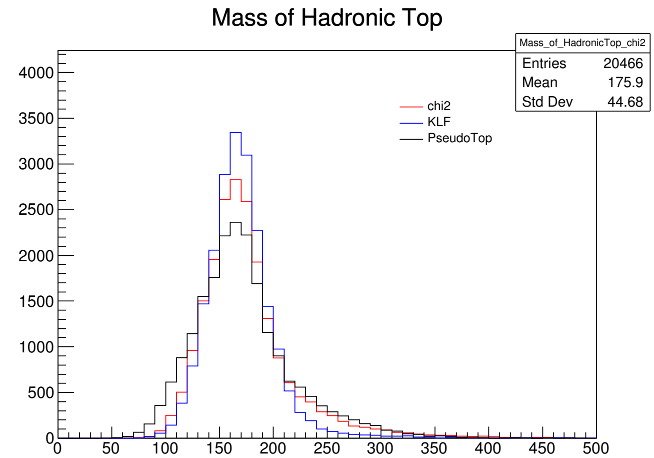
   chi^2 method considers all possible jet permutations and keeps the one with the lowest χ^2 value. The choice was made based on the equality of the tops' masses and the difference between the m_{jj} and the well-known value of m_W, taking into account the experimental resolution. As well as Pseudo Top approach, this method follows the exact same path for the W_{lep} reconstruction.

3. KLFitter [4] First, this reconstruction technique is a likelihood-based method and is only applicable to unfolding data to parton level. What is trying to do is to fit the p4 of each decay product taking into account jets' and MET's resolution. Subsequently, the fitted p4 information was ignored, since this was used just to identify the correct jets

The goal of these reconstruction methods is to correctly identify the reconstructed jets with the tops' decay products and to achieve a descent measurement of the top kinematics. In order to fairly compare the 3 methods, we fix these cuts to give us the same efficiency of well-reconstructed events.

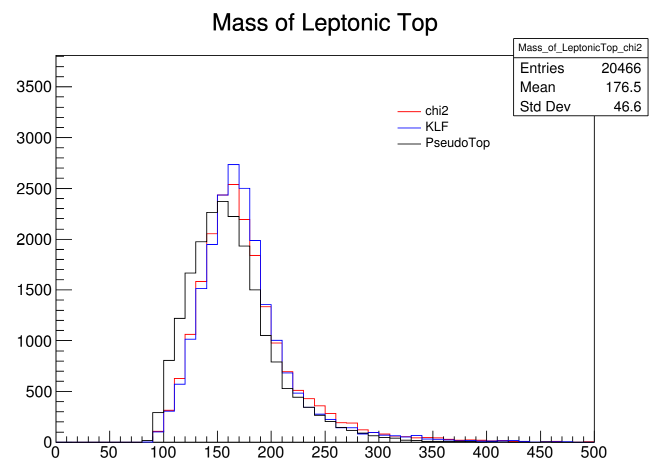
- χ^2 cut : $\chi^2 < 30$
- Pseudo Top cut : $\chi^2 < 60$
- KLFitter cut : LLH > -58

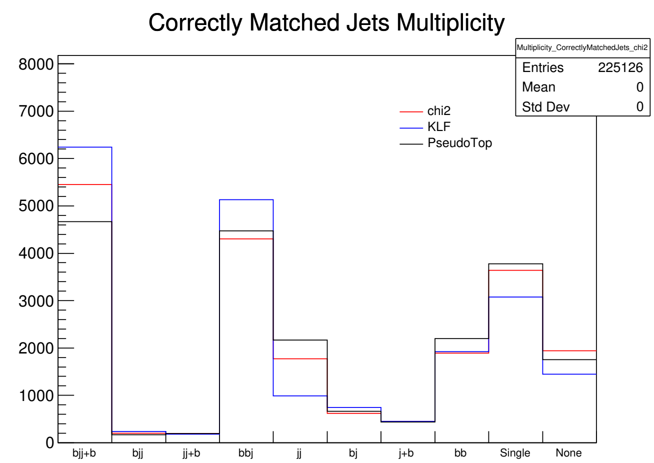




(a) *W_{had} mass* ***(b)*** *top_{had} mass*

Figure 3.1





(a) *top_{lep} mass* ***(b)*** *Correctly matched jets with partons*

Figure 3.2

At first glance the PseudoTop approach seems to be doing an excellent job (Fig.3.1a), giving the narrowest and higher pick mass distribution for the W_{had}. But having a look at Fig.3.1b, the behaviour is completely inverted! Despite the fact that KLFitter seems to make the worst choice for the W_{had} candidate, at the end it gives the best top_{had} (and top_{lep}, Fig.3.2a) distribution. Taking into account and the different sorting of reconstruction methods for the 1^{st} and 5^{th} bin on Fig.3.2b (1^{st} bin corresponds to cases where all 4 jets were correctly matched with the true partons and the 5^{th} one to cases were only the 2 light jets were correctly matched), is easy to understand on what each method focuses on. PsuedoTop approach is trying to do the best guess for the W_{had}, resulting to a penalized top_{had} reconstruction. In addition, KLFitter (to some extent, chi^2

as well) focuses more on as good as possible t_{lep} & t_{had} reconstruction rather than a descent W_{had} and as it seems, is more efficient in total.

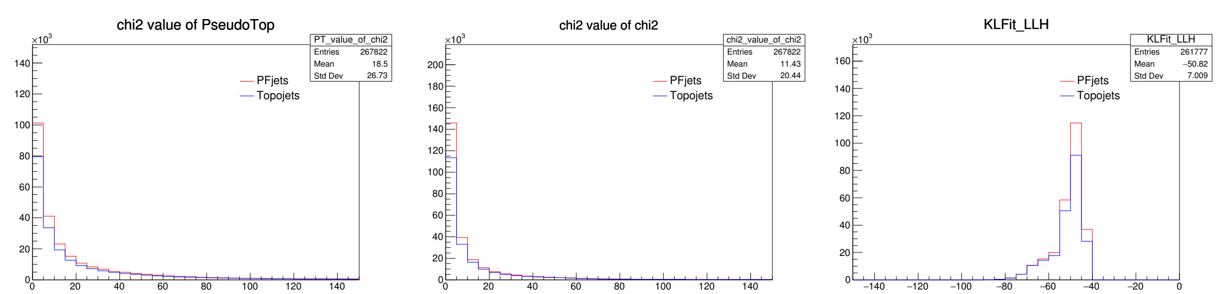
4 Topological Clusters Vs Particle Flow Algorithms

As a final part of this report, we are going to investigate if the accuracy of $t\bar{t}$ reconstruction is strongly depends on the jets' reconstruction accuracy. In doing so, 2 different jet reconstruction algorithms were used, the Topological clustering algorithm and the ParticleFlow algorithm. The major difference between the 2, is that in addition of Topological clusters algorithm [5](Sec.5) which is trying to suppress noise & reconstruct jets based on calorimetric information, Particle Flow algorithm[5](Sec.6) adds the tracking information. As a result, PF algorithm is expected to give better measurements for (mostly soft) charged hadrons.

Starting with almost the same number of total events, (PF with 1.03 M and Topo with 0.97 M), from both the distributions on Fig.4.1 and the 2 last columns of Table.4.1, regardless the reconstruction technique, by using PFjets instead of Topojets, the efficiency for well-reconstructed events, has been increased by $\sim 25\%$.

Total(evetns)		no-cuts		after-cuts	
PF	Topo	PF	Topo	PF	Topo
1.03M	0.97 M			increase by $\sim 25\%$	
PseudoTop		267822	222623	227790	185419
χ^2		267822	222623	228182	184684
KLF		261777	218033	223625	181847

Table 4.1



(a) ***(b)*** ***(c)***

Figure 4.1

Having a look at the t_{had} and W_{had} mass distributions comparison plots (Fig.4.2), which have been normalized to unity for fair comparison, they are almost similar but still, for PFjets are marginally better and narrower than for Topojets.

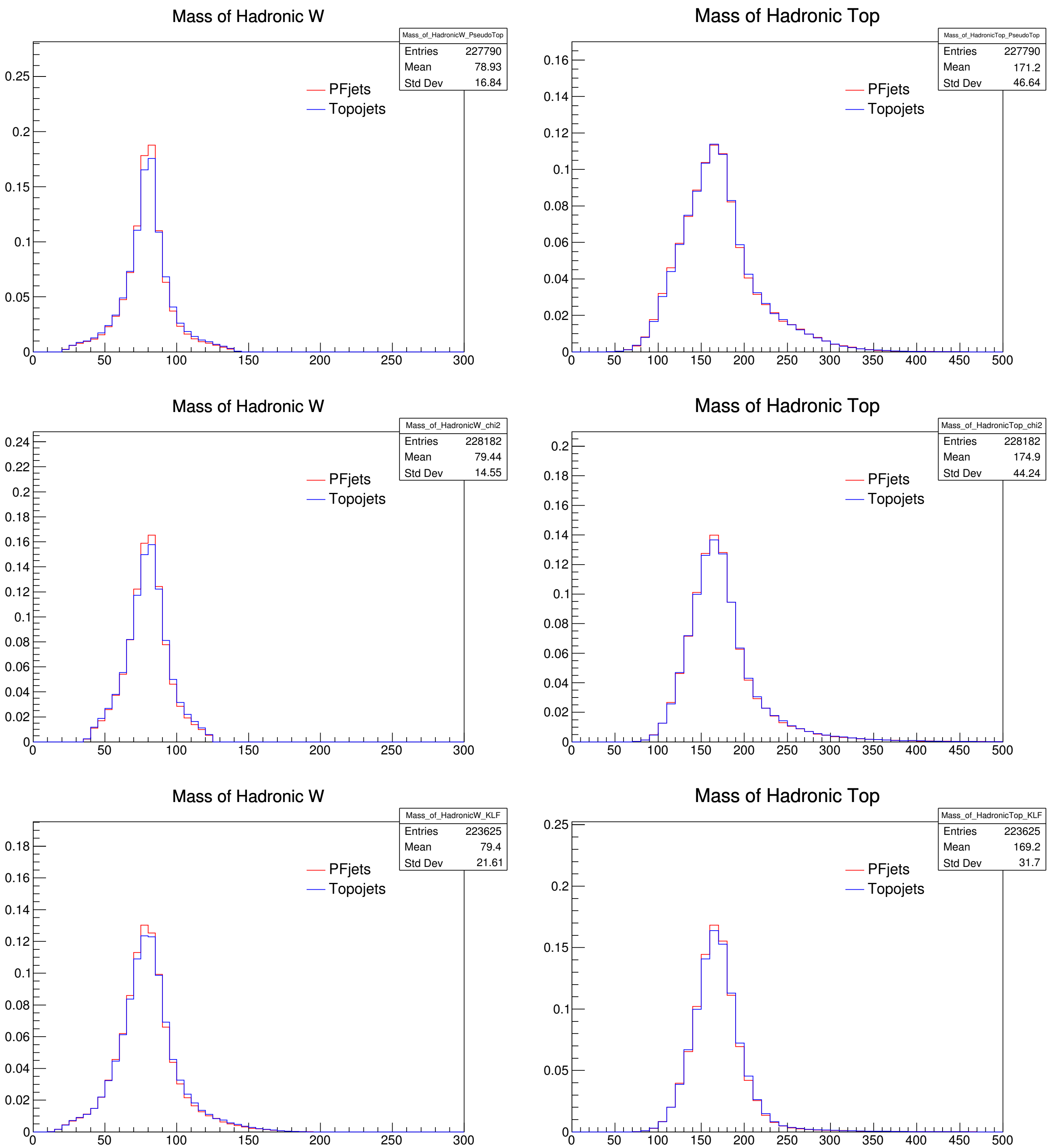


Figure 4.2: *Mass distributions using Topo and PF jets*

5 Conclusion and Outlook

The semileptonic $t\bar{t}$ channel could be challenging indeed. Overestimation of MET seems to lead to unphysical results but even for the realistic-like cases, the ΔR between the real neutrino and the reconstructed one is not satisfying. That's why the pursuit of a better MET guess is mandatory. From the 3 reconstruction methods comparison, the major deduction is that by optimizing the reconstruction of W_{had}, the reconstruction of t_{had} penalized, that;s why PseudoTop picks the dijet and gives the best W_{had} candidate and at the same time the worst top_{had} (and top_{lep}). In contrast, KLFitter focuses on choosing the best top_{had} and top_{lep} candidates, no matter the cost on making a good guess for the W_{had}. Concerning the comparison between the Topological clusters algorithm and the ParticleFlow algorithm, using ParticleFlow jets instead of Topo-jets, in the first place the reconstructed quantities have slightly better resolution, but the efficiency for succesful top-pair reconstruction could be increase by $\sim 25\%$ (see fig.5.1).

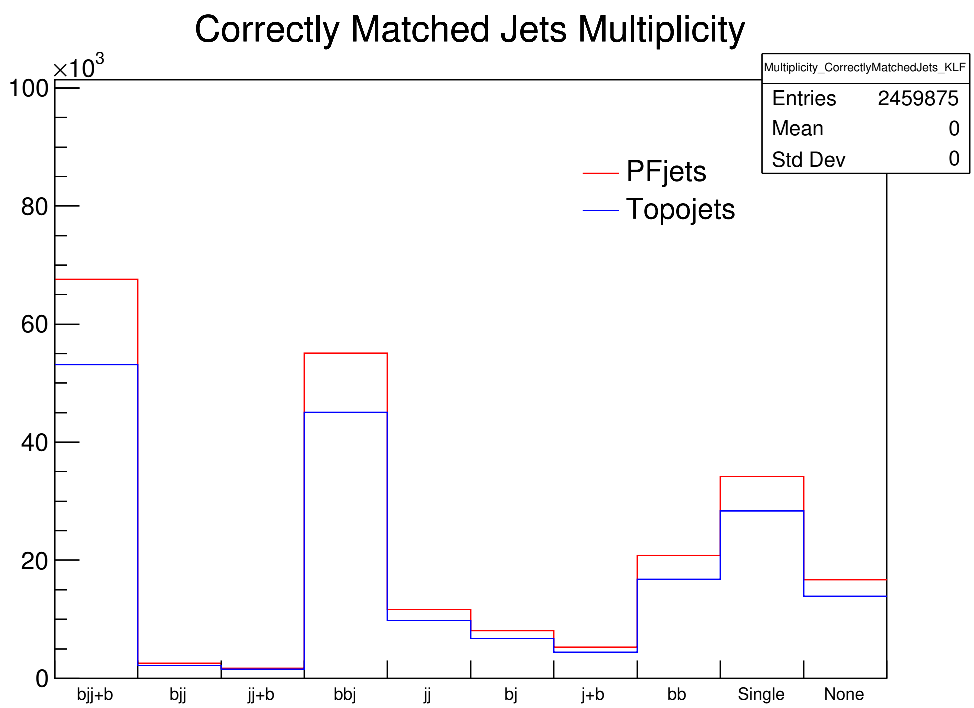


Figure 5.1: *Correctly matched jets with partons for KLFitter method, comparing the 2 different jet-reconstruction algorithms*

References

[1] *The Large Hadron Collider Project*, http://lhc.web.cern.ch/lhc/

[2] *ATLAS Detector*, https://atlas.cern/discover/detector

[3] ATLAS Collaboration, *Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP09 (2017) 118, doi:10.1007/JHEP09(2017)118, arXiv:1702.07546.

[4] Johannes Erdmann, Stefan Guindon, Kevin Kroeninger, Boris Lemmer, Olaf Nackenhorst, Arnulf Quadt, Philipp Stolte, *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nuclear Instruments and Methods in Physics Research Section A 748 (2014) 18-25, doi:10.1016/j.nima.2014.02.029 arXiv:1312.5595

[5] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C 77 (2017) 466, doi: 10.1140/epjc/s10052-017-5031-2 arXiv:1703.10485