

December 21, 2018

Reconstruction of  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  events  
using the matrix element method  
and substructure techniques

MAREN MEINHARD

*Institute for Particle Physics and Astrophysics*  
*Eidgenössische Technische Hochschule Zürich, 8093 Zürich, SWITZERLAND*

This contribution outlines the implementation of the matrix element method (MEM) in the search for  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  events. In particular, the evaluation of the transfer functions, which relate detector level to parton level quantities in the computation of the MEM, is described. In addition, it is presented how jet substructure reconstruction can be combined with the MEM. The combination of these techniques leads to a decrease in computation time by up to 90% and an increase in event selection efficiency of 30% in the high Higgs boson and top quark  $p_T$  phase space.

PRESENTED AT

11<sup>th</sup> International Workshop on Top Quark Physics  
Bad Neuenahr, Germany, September 16–21, 2018

# 1 Introduction

In the absence of direct observation of beyond the standard model (BSM) physics at the LHC, precision measurements are becoming the main pathway to test the consistency of the standard model (SM). In this context the associated production of a Higgs boson with a top-antitop quark pair ( $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$ ) is particularly interesting, since this channel allows for a direct measurement of the top and bottom quark Yukawa couplings. Both the  $t\bar{t}H$  production process and the Higgs boson decay to bottom quarks have been observed by the CMS and ATLAS collaborations in 2018 [1]. The identification of  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  events is challenging due to the complicated final state involving many jets, leptons and MET, and by the irreducible background  $t\bar{t}+b\bar{b}$  whose cross section ( $\sim 832$  pb) is over three magnitudes larger than the  $t\bar{t}H$  cross section ( $\sim 0.5$  pb). Furthermore the combinatorial self-background for various possibilities to associate jets to initial quarks further complicates the matter, as well as the large theoretical uncertainties on the predictions for the  $t\bar{t}+b\bar{b}$  process.

In order to overcome these difficulties, dedicated analysis techniques are necessary. In particular, the matrix element method is well suited for this analysis channel since it intrinsically solves the combinatorial background issue.

## 2 The matrix element method

The matrix element method (MEM) [2] is a fully analytical method which computes\* the probability for an event with characteristics  $\mathbf{y}$  (kinematics of jets, MET, leptons...) to originate from the underlying process  $\alpha$ :

$$P(\mathbf{y}|\alpha) \propto \frac{1}{\sigma_\alpha} \int d\Phi(\mathbf{x}) |M_\alpha|^2(\mathbf{x}) W(\mathbf{x}, \mathbf{y}) \quad (1)$$

where  $\sigma_\alpha$  is the total cross section of process  $\alpha$ ,  $d\Phi(\mathbf{x})$  is the phase space measure,  $|M_\alpha|^2(\mathbf{x})$  is the LO scattering amplitude squared and  $W(\mathbf{x}, \mathbf{y})$  represents the probability to obtain a detector response  $\mathbf{y}$  for a particle level event  $\mathbf{x}$ . A discriminant can be obtained by taking the ratio of the signal  $P(\mathbf{y}|t\bar{t}H, H \rightarrow b\bar{b})$  and the background  $P(\mathbf{y}|t\bar{t} + b\bar{b})$  probabilities.

Since the parton level quantities  $\mathbf{x}$  are unknown, they must be related to the observed quantities  $\mathbf{y}$  via transfer functions  $W(\mathbf{x}, \mathbf{y})$ . These are evaluated separately for each type of particle in the detector. Considering the much better momentum and angular resolution for leptons than for jets, a Dirac delta function models the transfer functions for leptons. For the MET, a multivariate normal distribution is used:

---

\*The CUBA package [3] is used for numerical integration.

$$W_{\text{MET}}(\mathbf{p}_T | \sum_k \mathbf{p}_k) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{p}_T - \sum_k \mathbf{p}_k)^T \Sigma^{-1} (\mathbf{p}_T - \sum_k \mathbf{p}_k) \right] \quad (2)$$

with  $\Sigma = \sigma_{\text{MET}} \mathbf{I}$  and  $\sigma_{\text{MET}} = 30$  GeV.

Transfer functions for jets are evaluated from MC simulation, taking into account the varying detector resolution according to a jet's  $p_T$ , pseudorapidity and flavour. A double gaussian distribution models the jet  $p_T$  distribution corresponding to quarks within a specific  $p_{T,\text{gen}}$  as shown in Equation 3 and Figure 1 (left). An across-bin fit in  $p_T$  is performed for each of the four parameters of the double-gaussian fit ( $\alpha_i$  for  $i \in 1, \dots, 4$ ). The jet transfer functions can then be reconstructed using these across-bin fits, and it can be seen in Figure 1 (left) that this reconstruction models the distribution well.

$$TF(p_T | p_{T,\text{gen}}) = N \left[ 0.7 \exp \left( \frac{p_{T,\text{gen}} - p_T - \alpha_1}{\alpha_2} \right)^2 + 0.3 \exp \left( \frac{p_{T,\text{gen}} - p_T - \alpha_3}{\alpha_2 + \alpha_4} \right)^2 \right] \quad (3)$$

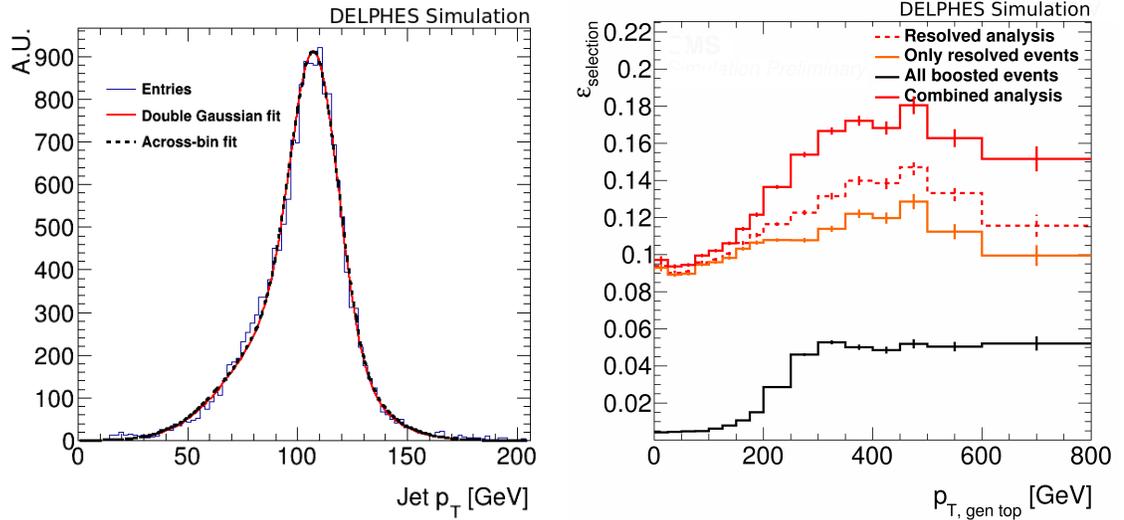


Figure 1: Double-gaussian and across-bin fits of jets originating from quarks with  $p_{T,\text{gen}} \in [100,102]$  GeV (left). Event selection efficiency as a function of generated top quark  $p_T$  for standard resolved analysis (red, dashed), resolved only (orange), events with HTTV2 candidate (black) and combined boosted analysis (red, solid) (right).

In evaluating the MEM probability, all possibilities to associate the jets to the underlying quarks need to be considered. In particular for  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  events, 12

permutations are considered to map all four b tagged jets to the b quarks produced by the decay of the Higgs boson and both top quarks. In order to perform these permutations, at least four b tagged jets are required for the MEM computation.

### 3 Merging substructure techniques with the MEM

When the top quark and/or the Higgs boson in  $t\bar{t}H$ ,  $H\rightarrow b\bar{b}$  events have an increasingly high transverse momentum, their decay products become collimated and might merge into a single classical jet, therefore failing the mapping of the observed jets to the quarks in the evaluation of the MEM. This can be solved using substructure techniques. In these proceedings, emphasis will be put on tagging top quarks with high  $p_T$ . However, tagging Higgs bosons is an equally promising idea. Furthermore, only events with exactly one top quark decaying leptonically are considered.

Signal and background ( $t\bar{t}$ +jets) events are generated using POWHEG v.2 [4] while PYTHIA v.8.200 [5] is used to model hadronization and parton shower. The detector response is modelled with DELPHES [6]. All other specifications such as pileup contributions correspond to the 2017 LHC conditions. Jets are reconstructed with the anti- $k_T$  algorithm [7] ( $R=0.4$ ) and tagged as b quarks if they pass the loose working point (corresponding to a  $\sim 10\%$  light-flavour misidentification rate).

The HEPTopTagger V2 algorithm (HTTV2) [8] is used for top quark identification. Using as input Cambridge-Aachen [9] large- $R$  jets with a distance parameter of 1.5, the algorithm reconstructs exactly three subjets from the large- $R$  jet, while removing contributions from pileup and underlying events.

Figure 2 illustrates how information from HTTV2 candidates can be used in the MEM analysis. To do so, new input lists of jets for the MEM are created, using one b tagged subjet and two light tagged subjets from the HTTV2 candidate. Resolved b tagged jets which cannot be associated to HTTV2 subjets are added to the new input list of b jets, which is then used for the calculation of the boosted MEM. Hence resolved jets are replaced with HTTV2 subjets in the boosted MEM. For events with no HTTV2 candidate, the standard, resolved MEM can be computed.

Combining substructure techniques with the MEM leads to several improvements. In particular the long computation time for the MEM evaluation, which is one of its major drawbacks especially when looking ahead at HL-LHC conditions, can be significantly reduced for events with a HTTV2 candidate. In the standard resolved analysis, the computation of the MEM requires  $52.5 \pm 27.4$  s<sup>†</sup> per event. When replacing jets with HTTV2 subjets, the computation time is reduced by  $\sim 75\%$  to  $13.7 \pm 1.4$  s. This is because many events have additional jets coming from QCD radiation. Knowing which non b tagged jets arise from the top quark decay, QCD

---

<sup>†</sup>Evaluated using standard x86 CPU units with a Intel Xeon X5560 processor. The uncertainties represent the standard deviation.

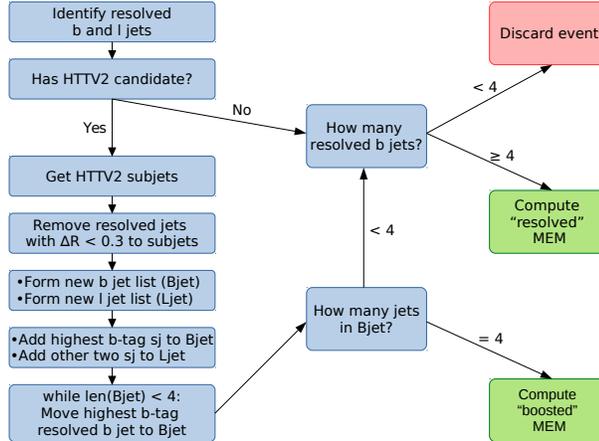


Figure 2: Analysis strategy for merging top quark candidates with the MEM.

jets can be removed, thus reducing the number of permutations to consider.

The computation time can be further reduced as the HTTV2 subjets correspond to the top quark’s decay products. Therefore, only 3 permutations need to be considered to map all jets to the generator level quarks. In this case, the computation time is reduced to  $3.5 \pm 0.3$  s, which corresponds to a total speed-up of 90%.

Furthermore, the newly devised analysis strategy allows to recover events for which the MEM could not be evaluated with resolved jets. This occurs when too few b tagged jets are identified, which happens when a top quark has a large  $p_T$ , and the b quark cannot be reconstructed as a single resolved jet. These events can be recovered by reconstructing the top quark’s decay structure with the HTTV2 algorithm. Figure 1 (right) shows the event selection efficiency as a function of the generated top quark  $p_T$  for the standard resolved and the combined boosted analysis. As expected, the efficiency in the latter one exceeds the selection efficiency of the resolved analysis in the regime of top quarks with large  $p_T$  by up to 30%. Since the  $p_T$  of the top quarks and the Higgs boson in the  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  channel are correlated, this also leads to an increased event selection efficiency for Higgs bosons with large  $p_T$ .

This is promising since many BSM theories predict deviations from the SM in the high  $p_T$  region of the phase space. For example, deviations from the SM CP structure of the top quark Yukawa coupling [10] could appear at large Higgs boson  $p_T$  or additional terms in the standard model Lagrangian [11] could lead to a modified momentum distribution of the Higgs boson. Identifying events in this phase space would therefore enable to test these theories.

## 4 Conclusion

The matrix element method is a powerful analysis technique which allows the reconstruction of complicated final states such as the  $t\bar{t}H$ ,  $H \rightarrow b\bar{b}$  process. A key part of the MEM computation is the evaluation of the transfer functions. It has been shown that they are well modelled by double gaussian distributions. Furthermore, a strategy to combine substructure methods, in particular top quark identification, with the MEM has been presented. A major advantage of this method is the reduction of the computation time by up to  $\sim 90\%$ . This comes from the reduction of the number of permutations, being able to identify QCD radiation jets, and removing part of the self-background. Furthermore, the event selection efficiency can be enhanced in the regime of top quarks and Higgs bosons with large  $p_T$ , which is crucial for testing various BSM theories which affect this region of the phase space.

## ACKNOWLEDGEMENTS

I am grateful to A. Gomez, G. Kasieczka, J. Pata and C. Reifel for their support and constructive comments which were essential to the realization of this work.

## References

- [1] ATLAS Collaboration, Phys. Lett. B **784**, 173 (2018). CMS Collaboration, Phys. Rev. Lett. **120**, 231801 (2018). ATLAS Collaboration, Submitted to PLB (2018). CMS Collaboration, Phys. Rev. Lett. **121**, 121801 (2018)
- [2] P. Artoisenet *et al.*, Phys. Rev. Lett. **111**, 091802 (2013). J. Pata, PhD Thesis (2018)
- [3] T. Hahn, Comput. Phys. Commun. **168** 78 (2005)
- [4] P. Nason, JHEP **11** 040 (2004)
- [5] T. Sjöstrand *et al.*, Comput. Phys. Commun. **191** 1583 (2015)
- [6] The DELPHES 3 collaboration *et al.*, JHEP **1402** 057 (2014)
- [7] M. Cacciari *et al.*, JHEP **04** 063 (2008)
- [8] G. Kasieczka *et al.*, JHEP **1506**, 203 (2015)
- [9] Y. L. Dokshitzer *et al.*, JHEP **9708**, 001 (1997)
- [10] M. Buckley & D. Gonçalves, Phys. Rev. Lett. **116**, 091801 (2016)
- [11] J. Bramante *et al.*, Phys. Rev. D **89**, 093006 (2014)