PROCEEDINGS OF SCIENCE

A Top Friendship: Measurement of $t\bar{t}H$ production in the $H(bb)$ decay channel at ATLAS with Transformer Networks

Yusong Tian[a,b,*] **on behalf of the ATLAS collaboration**

[a] *II. Physikalisches Institut, Georg-August-Universität Göttingen, Friedrich-Hund-Platz 1, DE - 37077 Göttingen, Germany*

[b] *now at Département de Physique Nucléaire et Corpusculaire, University of Geneva, Quai Ernest-Ansermet 24, 1211 Geneva, Switzerland*

E-mail: yusong.tian@cern.ch

The exploration of the Higgs boson's properties and its interactions with top quarks constitutes a pivotal aspect of the post-Higgs discovery era. Among these, the measurement of the associated production of a Higgs boson with a pair of top quarks ($t\bar{t}H$) offers a unique window into the Yukawa coupling between the Higgs and the top quark, the heaviest known fundamental particle. These proceedings present the latest results on $t\bar{t}H$ production with the Higgs boson decaying into a pair of bottom quarks ($H \rightarrow b\bar{b}$), performed by the ATLAS collaboration. A key focus is placed on the improved MVA strategy, which incorporates permutation-invariant architectures utilising attention mechanisms, for enhancements in both multi-class classification of signal and background events, and Higgs candidate reconstruction.

ATL-PHYS-PROC-2024-068
30/09/2024

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

*Speaker

1. Introduction

This $t\bar{t}H(b\bar{b})$ Run 2 legacy analysis [1] was done on 140 fb^{-1} of 13 TeV pp data collected by the ATLAS experiment at the Large Hadron Collider in one or two light lepton (electron/muon) final states. Events were separated into three channels: dilepton, single-lepton resolved, and single-lepton boosted (the boosted channel was selected by a deep neural network). The main challenge arises from the large $t\bar{t}$+jets background due to jets from b- or c- quarks. The signal strength $\mu = \frac{\sigma_{\text{measured}}}{\sigma_{\text{SM prediction}}}$ was measured in the simplified template cross-section (STXS) framework, to probe Higgs transverse momentum (p_{T}^{H})-dependent deviations from the Standard Model (SM): Events get separated into signal and control regions (SR and CR) with different STXS p_{T}^{H} bins, as shown in Figure 1.
The analysis benefited from various improvements with respect to the previous version [2], including transformer neural networks (TNN) used for multivariate analysis. Two TNNs were used for event classification and p_{T}^{H} reconstruction, allowing the separation of the events into SR and CRs, and also into different STXS regions.

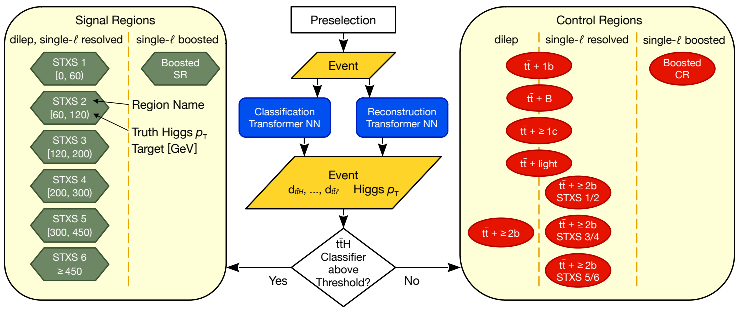


Figure 1: Flowchart of separating events into different regions.

2. Event Classification and Higgs p_{T} Reconstruction

Firstly, the events go through a preselection based on the number of jets with radius $R = 0.4$ or $R = 1.0$ in the event, and the number of b-jets in them. The selection criteria are looser than that in the previous analysis [2], resulting in a 6.3 % preselection efficiency for $t\bar{t}H$ signal events - around three times higher than before.
The events then go through two TNNs. Through the Classification TNN, each event obtains six discriminant values for each class: $t\bar{t}H$ signal, and $t\bar{t}+1b$, $t\bar{t}+1B$, $t\bar{t}+\geq 2b$, $t\bar{t}+\geq 1c$, $t\bar{t}$ + light backgrounds. The discriminant is defined as:

$$d_i = \frac{p_i}{\sum\limits_{j\neq i} p_j \cdot \hat{N}_{ij}},\ \hat{N}_{ij} = \frac{N_j}{\sum\limits_{k\neq i} N_k}, \tag{1}$$

where i, j, and k denotes the classes, p_i is the probability that this event is in class i, N is the expected event yield, and $\hat{N}_{ij}$ are the weights. This discriminant defined provides good performance, and is used for event classification in signal/control regions as well as discriminant in the profile-likelihood fit.
The cut values on $d_{t\bar{t}H}$ above which an event is considered signal-like were determined by $\max(\frac{\text{Signal}}{\sqrt{\text{Background}}})$, and the background grouping by the highest d_i for background classes.
The Reconstruction TNN predicts which two jets are most likely from the H decay, and p_{T}^{H} is calculated from their four-

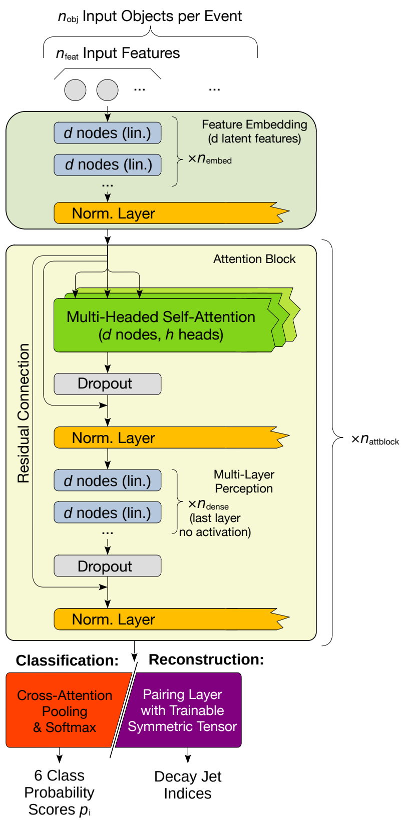


Figure 2: The transformer model architecture.

vectors. This method gave better performance than obtaining p_{T}^{H} from regression, the predictions of which shift to central STXS bins.

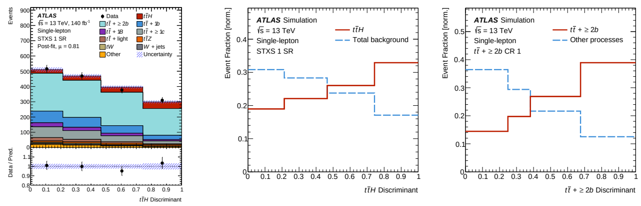
3. The Transformer Neural Networks

Transformer models like GPT [3] utilise the attention mechanism [4]. A similar architecture, as shown in Figure 2, was used in this analysis. The architecture is permutation-invariant, which means the result is independent of the ordering of the input objects. The input features are kinematic information of reconstructed electrons, muons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$ and the b-tagging score. The classification TNN was trained on $t\bar{t}H$ and $t\bar{t}$ + jets events, and the reconstruction TNN only on $t\bar{t}H$ signal events.

The residual scheme is different from that in Ref. [4], where two separate residual connections are employed around the multi-head self-attention and the multi-layer perception layers. The modified scheme results in less dependency on earlier attention blocks and gives a slightly higher average classification accuracy. For the reconstruction TNN, the pairing layer is similar to that in Ref. [5].

4. Performance

The performance of the classification TNN is illustrated in Figure 3. Figure 3(a) shows that the distribution of background events decrease with $d_{t\bar{t}H}$ while signal events increase, and good data/MC agreement. Figure 3(b) and 3(c) shows $t\bar{t}H$ and $t\bar{t}+\geq 2b$ classes as examples: the distribution of the target class peak at higher values of the corresponding discriminant values $d_{t\bar{t}H}$ and $d_{t\bar{t}+\geq 2b}$, showing good performance in event separation. The performance on other background classes is similar. The TNN approach in this analysis improved over the BDT approach as used previously in Ref. [6]. The area under the ROC curve is 0.753 for the single-lepton resolved channel and 0.774 for the dilepton channel.



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

Figure 3: (a): post-fit separation of event classes on MC compared to data in the single-lepton channel. (b), (c): pre-fit separation of $t\bar{t}H$, $t\bar{t}+\geq 2b$ events in single-lepton resolved channel in MC versus their respective discriminant values. The discriminant values were rescaled to be between 0 and 1.

5. Results of the Analysis

Combined profile-likelihood fit found an excess of $t\bar{t}H(b\bar{b})$ events with an observed (expected) significance of 4.6 (5.4) standard deviations. The $t\bar{t}H$ signal strength is $\mu_{t\bar{t}H} = 0.81 \pm 0.11(\mathrm{stat.})\,^{+0.20}_{-0.16}(\mathrm{syst.})$ for $m_H = 125.09\,\mathrm{GeV}$, with the cross-section $\sigma_{t\bar{t}H} = 411 \pm 54(\mathrm{stat.})\,^{+85}_{-75}(\mathrm{syst.})$ fb. The signal strengths were also measured in STXS bins. All results are consistent with the SM predictions.

References

[1] ATLAS Collaboration, (2024), arXiv: 2407.10904 [hep-ex]

[2] ATLAS Collaboration, JHEP 06 (2022) 097

[3] T. Brown et al., *Advances in Neural Information Processing Systems*, **33**, 2020, p. 1877

[4] A. Vaswani et al., *Advances in Neural Information Processing Systems*, **30**, 2017, p. 5998

[5] M. J. Fenton et al., Phys. Rev. D 105 (2022) 112008

[6] ATLAS Collaboration, Phys. Rev. D 97 (2018) 072016