Boosting the Higgs with $t\bar{t}H(b\bar{b})$

Using the ATLAS detector

Albert Borbely
albert.borbely@cern.ch

University of Glasgow

March 21, 2022

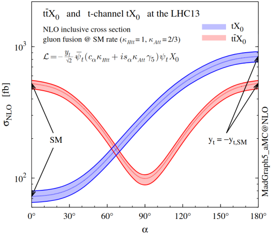
Table of Contents

University of Glasgow

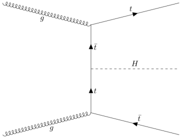
ATLAS EXPERIMENT

Introduction

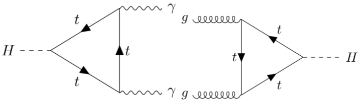
- The Higgs-top Yukawa coupling (y_t) is one of the parameters of the Standard Model (SM) Lagrangian.
- It can be accessed directly with $t\bar{t}H$ (tH) or indirectly through a virtual top quark loop e.g. with $H \to \gamma\gamma$ decays or ggF.
- $t\bar{t}H$ is **not** sensitive to the sign of y_t, tH **is** sensitive due to interference between W and top couplings.
- The $t\bar{t}H$ process has been observed at the LHC by both the ATLAS and CMS collaborations.
- The Higgs boson has many decay modes with its dominant decay mode being $H \to b\bar{b}$ (58%).



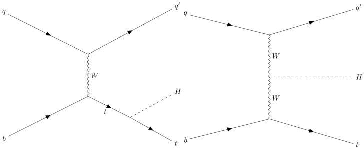
EUR. Phys. J. C 75, 267 (2015)



Direct: $t\bar{t}H$



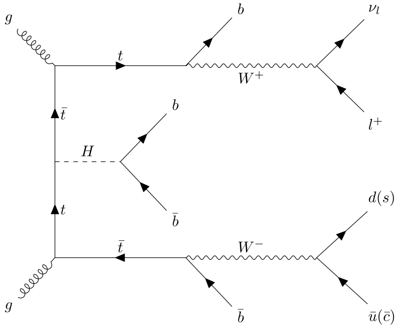
Indirect: $H \to \gamma\gamma$, ggF



Direct: tH

$t\bar{t}H(\bar{b}b)$ Overview

- The $t\bar{t}H$ process was initially observed by both the ATLAS and CMS collaborations in 2018.
  - Initial observation was done with the combination of many Higgs decay processes.
  - Driven by $H \rightarrow \gamma\gamma$
- Since then measurements of the $t\bar{t}H(\bar{b}b)$ process, using the full run-2 ATLAS data set ($139 fb^{-1}$), has been published (arxiv).
- Some key features of this analysis are:
  - First differential measurement in Higgs p_T using the (STXS 1.2) framework.
  - It is the combination of three signal regions.
  - Using single and di-lepton events.
  - Using resolved and boosted topologies.
- $t\bar{t}$ + *jets* dominant background.
- Work has begun on updating this measurement (Legacy analysis).
- Key updates:
  - Updated jet collections used
  - Updated *b*-tagging algorithm
  - Updating the various analysis specific machine learning tools



$t\bar{t}H$ decaying semileptonically.

- $t\bar{t}H$ represents $>$ 1% of Higgs production.

University of Glasgow

ATLAS EXPERIMENT

Signal regions

- The analysis is split into three main channels.
- Each channel uses different analysis techniques and is combined in a profile likelihood fit.
- Dilepton, where both tops decay leptonically, and the hadronic Higgs decay is resolved.
- Lepton + jets, where one of the tops decays leptonically, all jets are resolved.
- Boosted, where one of the tops decay leptonically, contains $\geq$ 1 boosted Higgs candidate.
- Each signal region can contribute to each of the bins used for the differential Higgs measurement.
- Events further split by jet and b-jet multiplicity into control and signal regions.
- Control regions (CR), have a low signal fraction, are used to constrain the backgrounds.

Boosted Topologies

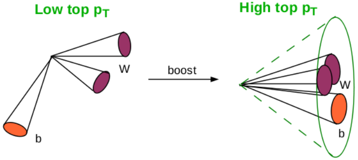


Figure: A boosted top decay.

- Decay products in the boosted "regime" have a high p_T so their decay products are more collimated.
- To deal with this a large anti-k_t jet, with $r = 1.0$, is used.
- The Higgs and hadronic top candidates are reconstructed this way.
  - Boosted Higgs candidates $p_T \geq 300$ GeV.
- Large jets can have substructure variables associated with them.
  - Useful for jet identification.

Background modelling

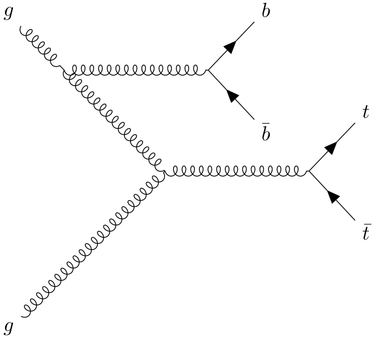


Figure: $t\bar{t} + b\bar{b}$ decay.

- Dominant background $t\bar{t}+$ jets
- Identical topology
- Has the same $4b$-jet final state
- Modelling uncertainties are needed as it can be simulated with both 5 and 4 flavour schemes.
- 4 flavour scheme is used for the nominal sample.

Recent Result

Recent result

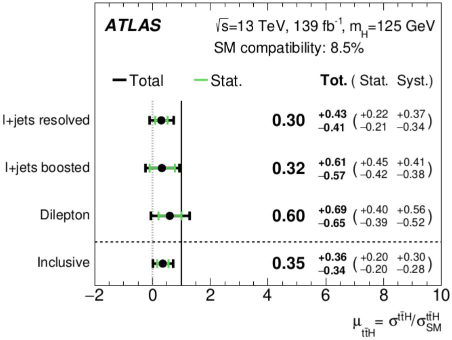


Figure: arXiv 2111.06712

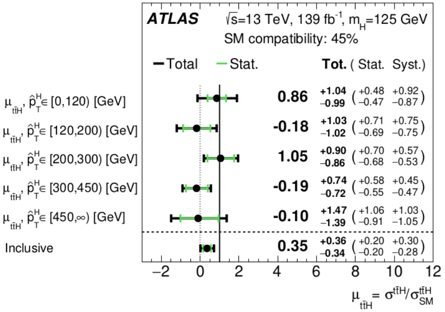


Figure: arXiv 2111.06712

- Measured signal strength $0.35^{+0.36}_{-0.34}$
- Observed(Expected) $= 1.0(2.7)\sigma$
- Probability the signal strength compatible with SM prediction 8.5%.

- $t\bar{t}+\geq 1b$ modelling dominant systematic uncertainty
- Boosted channel has a significant contribution.
- First differential measurement of $t\bar{t}H$ was performed in 5 bins of Higgs p_T in the STXS framework.

University of Glasgow

ATLAS EXPERIMENT

Recent results, Systematics

- The analyses dominated by $t\bar{t}+\geq 1b$ modelling systematics.
  - Largest pull on $t\bar{t}+\geq 1b$ ISR, 1.2σ
  - Large pull on $t\bar{t}+\geq 1b$: p_T^{bb} shape uncertainty
  - Large pull on $t\bar{t}+\geq 1b$: NLO match dilep. p_T^H
- This has been the primary challenge
- See systematics ranking plot on the right

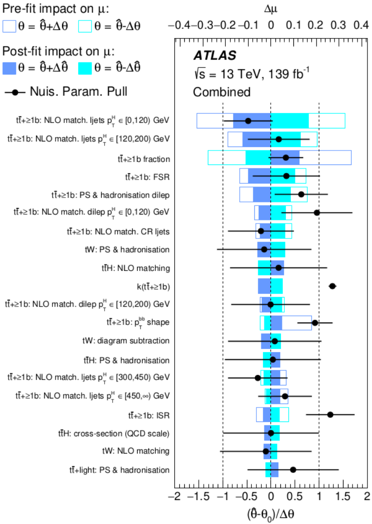


Figure: arXiv 2111.06712

University of Glasgow

ATLAS EXPERIMENT

Legacy Analysis

Jet objects

- The Legacy analysis will use updated jets.
- The initial result uses “EMtopo” jets which are topological clusters in the calorimeter.
- The anti-k_t algorithm is used to cluster the jets, with $r = 0.4$.
- The legacy analysis uses particle flow (pFlow) jets.
- It matches tracks to calorimeter clusters, to avoid double counting the energy of jets.
- Reduces the effects of pile-up, particularly at low p_T.

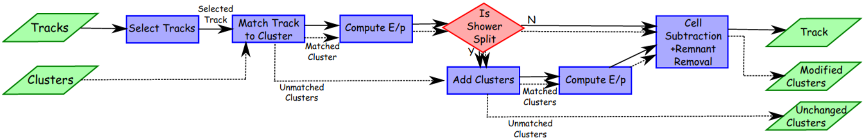


Figure: Particle flow algorithm, arXiv 1703.10485

University of Glasgow

ATLAS EXPERIMENT

b-tagging

- The analysis is highly dependent on b-tagging performance
- Final state contains 4 b-quarks
- With the l+jets/boosted regions having 6 jets and dilepton having 4
- Initial result used a BDT (MV2c10) taking in high level inputs
- The legacy analysis will be using an RNN (DL1r)

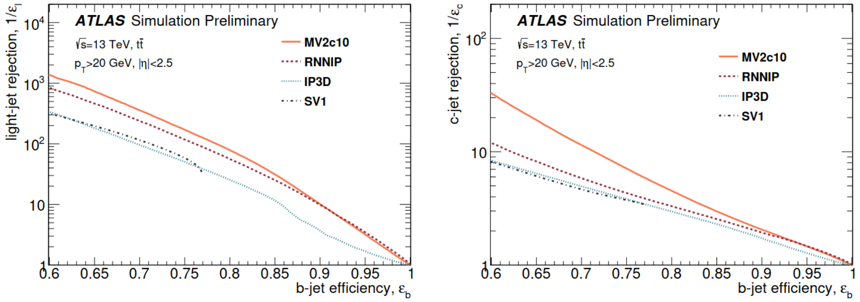


Figure: b-tagging performance DL1r (ROC), *light*-jet (left) and c-jet (right) rejection versus b-tagging efficiency, CDS link

University of Glasgow

ATLAS EXPERIMENT

Updating the boosted region

- In the previous analysis a DNN classifier was used for Higgs tagging boosted jets.
- This is currently being updated to use the new jet collections and *b*-tagging algorithm.
- Previously re-clustered jets (RC-jets) were used for the boosted region.
  - They are clustered from jets ($r = 0.4$) with the anti-k_t algorithm and $r = 1.0$
  - Contain sub-jets due to re-clustering, provides *b*-tagging information
  - Provide a direct way to combine with resolved channels as systematics propagate over
- Large-R jets (LR-jets) are being investigated as an alternative
  - They are directly clustered from the calorimeter, using anti-k_t, with $r = 1.0$
  - Harder to combine with resolved channels as systematics don't directly propagate
  - Need to ΔR match jets to obtain "sub-jet" type information, (*b*-tagging)

- The DNN is trained on the nominal $t\bar{t}H$ MC sample
  - Trained on samples produced for this analysis
  - Limited computing power available
- Commonly DNNs would be fed only basic information, requires a lot of data
- As data is limited feature engineering is used to increase the performance
- DNN inputs (features) are created and their separation power for the classes are evaluated
- Extra substructure variables, namely τ_{32} and τ_{21} (*N*-subjettiness), were added
- Due to their modelling they require that the samples be produced with the full detector simulation, arXiv 1005.4568

University of Glasgow

ATLAS EXPERIMENT

Feature engineering

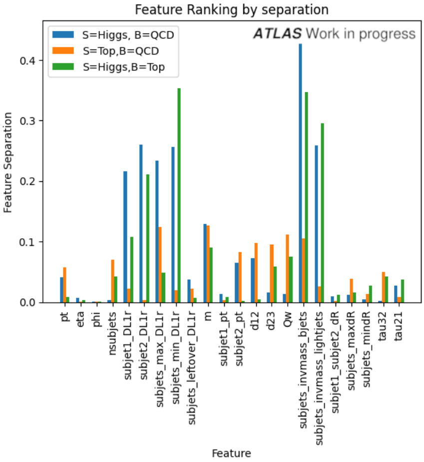


Figure: Separation power for DNN input features with LR-jets

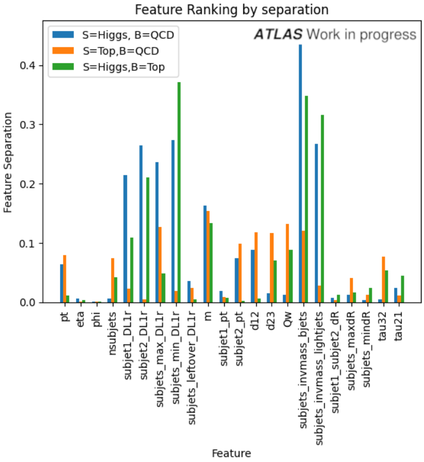


Figure: Separation power for DNN input features with RC-jets

DNN Boosted Higgs tagging RC-jets

- DNN performance is evaluated in a loose region (req. 1 RC-jet with 2 subjets and 4 jets in the event)
- Events with $\geq$ 1 Higgs-tag are selected
- $t\bar{t}H/t\bar{t}$ sample purity
- $t\bar{t}H$ frac is the number of remaining $t\bar{t}H$ events.

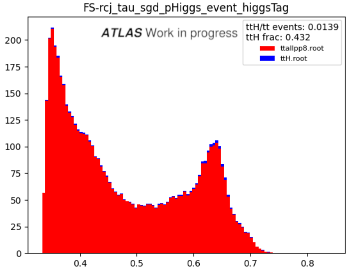


Figure: DNN output for Higgs probability with added τ substructure variables, trained with RC-jets and Full Sim. Samples scaled to luminosity.

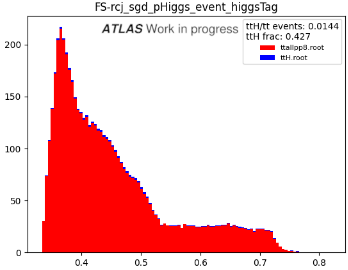


Figure: DNN output for Higgs probability, trained with RC-jets and Full Sim. Samples scaled to luminosity.

University of Glasgow

DNN Boosted Higgs tagging LR-jets

- DNN performance is evaluated in a loose region (req. 1 LR-jet with 2 ΔR matched jets and 4 jets in the event)
- Events with $\geq$ 1 Higgs-tag are selected
- $t\bar{t}H/t\bar{t}$ sample purity
- $t\bar{t}H$ frac is the number of remaining $t\bar{t}H$ events.

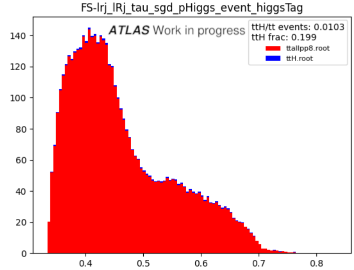


Figure: DNN output for Higgs probability with added τ substructure variables, trained with LR-jets and Full Sim. Samples scaled to luminosity.

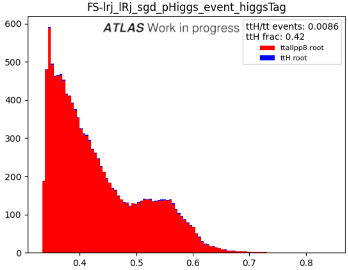
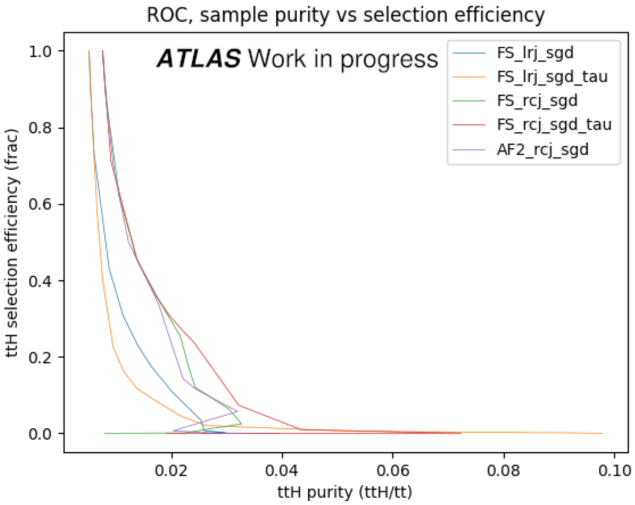


Figure: DNN output for Higgs probability, trained with LR-jets and Full Sim. Samples scaled to luminosity.

University of Glasgow

ROC curve


ROC, sample purity vs selection efficiency
ATLAS Work in progress
FS_lrj_sgd
FS_lrj_sgd_tau
FS_rcj_sgd
FS_rcj_sgd_tau
AF2_rcj_sgd
ttH selection efficiency (frac)
ttH purity (ttH/tt)
1.0
0.8
0.6
0.4
0.2
0.0
0.02
0.04
0.06
0.08
0.10

Conclusion

- There is no a significant improvement from adding in the extra τ (N-subjettiness) variables.
- There has also not been an improvement when using LR-jets.
- Both of these require full detector simulation.
- This would only be computationally feasible for the nominal samples.
- Systematic variations would still use fast simulation with simplified detector modelling.

Backup