

Searching for $H^+ \rightarrow W^+ h$ at ATLAS

Sid Baines

Under supervision of Ulla Blumenschein

QMUL

April 7, 2025



Overview

- 1 Motivation
- 2 Signal and backgrounds
- 3 Analysis Strategy
- 4 Event reconstruction
- 5 Event classification/Machine Learning
- 6 Alternative strategy: Parametrized Neural networks
- 7 Future ideas: Transformer-like networks
- 8 Results

Intro & motivation

- Search for $H^+ \rightarrow W^+ h \rightarrow (qq/l\nu)bb$
 - First search for this channel
- Motivated by BSM theories & complements previous searches (fig. 1)
 - Type-1 2HDM for $\cos(\beta - \alpha) \neq 0$
 - Georgi-Machacek model, 3HDM
- **Merged analysis**¹ is performed in tandem with a resolved search
- Paper published in JHEP [link]

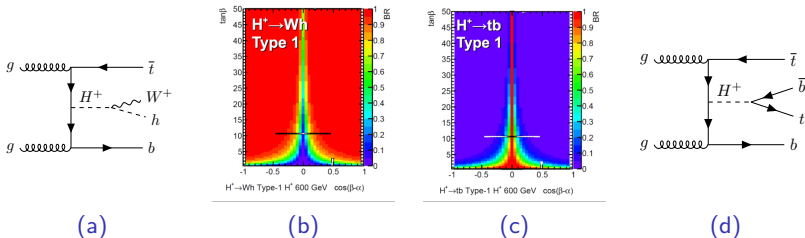


Figure 1: $H^+ \rightarrow W^+ h$ decay (a) has a branching ratio (b) complementary to the $H^+ \rightarrow tb$ decay (d) branching ratio (c).

¹ie with bb and/or qq treated as merged jet-pairs, rather than separate jets per quark

Signal & backgrounds

Signal

- $H^+ \rightarrow W^+ h$ produced with associated $\bar{t}b$ (see fig. 2)
- Charge agnostic (here refer to H^+ but assume either)
- Single lepton final state
 - Two decay modes (see fig. 2)
- Consider 10 mass points in range $m_{H^+} \in [0.8, 3.0]$ TeV

Backgrounds

- $t\bar{t}$ (dominant)
 - Enriched heavy flavour & high HT samples
- W^\pm/Z + jets, single top
- All modelled with MC - data-driven corrections investigated but not used

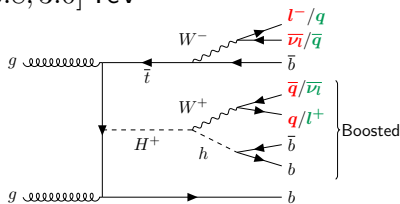


Figure 2: Representative Feynman diagram for the signal process

Analysis strategy

- Use full Run 2 dataset (140 fb^{-1})
- Reconstruct events by hand using large-R jets based on presence/properties of reconstructed objects
- Construct discriminating variables & train machine learning model to separate signal from background
- Split into Signal & Control regions based on channel, ML scores & b-tag multiplicity
- Statistical fit of invariant mass m_{W+h}

Event reconstruction/selection

- Single lepton, multi-jet final state with boosted quarks/leptonic-W
- Reconstruct SM Higgs as ' $X \rightarrow bb$ '-tagged large-R jet

$qqbb$

- **if** have another Large-R jet with $m_J \in [50, 110]$ GeV, use this as W^+ , ...

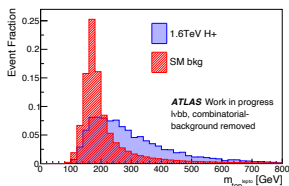
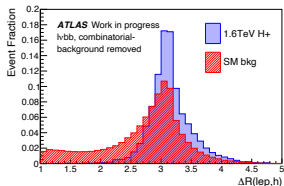
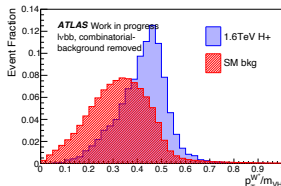
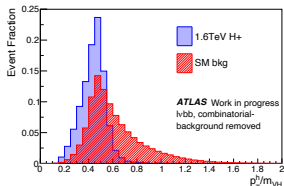
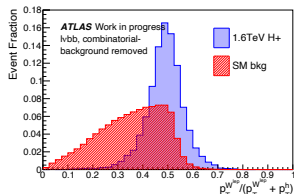
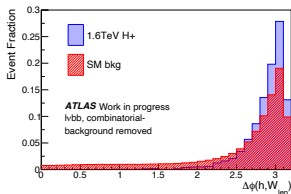
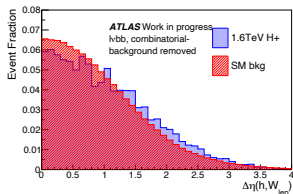
$lvbb$

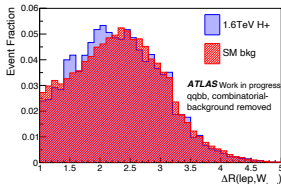
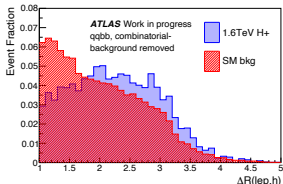
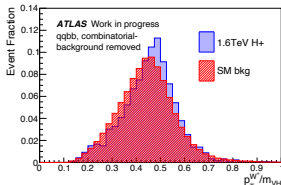
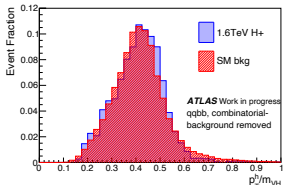
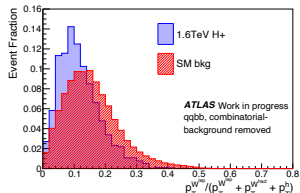
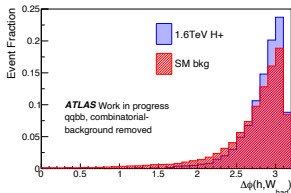
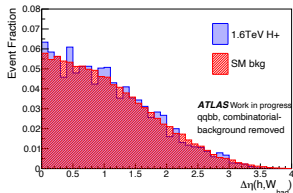
- ... **else** reconstruct W^+ using lepton & p_T^{miss}

- Count #b-tagged small-R jets outside of $H(/W^+)$ large-jet(s)
- Other reconstruction methods explored for recuperating loss signal/reducing combinatorial background

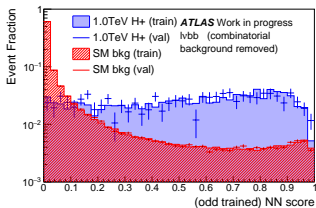
Machine learning for sig/bkg separation

- Reconstruction step: Keep 55-60% of signal, reduce major backgrounds by $\sim 92\%$
- But can reduce further with ML!
- Train separate neural networks (NNs) for each channel
 - Use 2-fold splitting per-channel so training data is never used for final prediction
- Inputs are kinematic variables such as angles, p_T -ratios, etc.
- Networks perform better at higher masses
 - Decision made to **train using $m_{H^+} > 1\text{TeV}$** to retain performance where merged channel most important

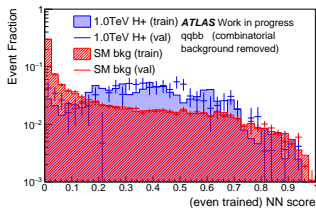
$lvbb$ NN inputs (sig vs 1.6TeV)

$qqbb$ NN inputs (sig vs 1.6TeV)

Machine learning discriminant

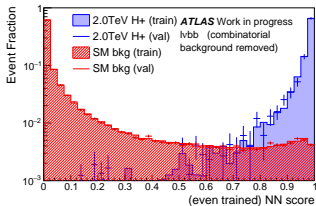


(a)

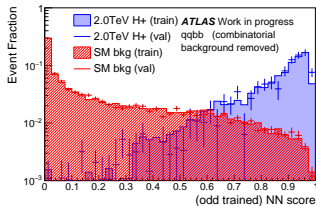


(b)

Figure 3: Distributions of background (red) and 1.0TeV signal (blue) samples for the $lvbb$ (a) and $qqbb$ (b) NNs



(a)

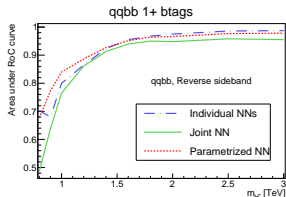


(b)

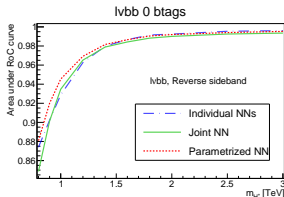
Figure 4: Distributions of background (red) and 2.0TeV signal (blue) samples for the $lvbb$ (a) and $qqbb$ (b) NNs

Alternative strategy: Parametrized Neural Networks

- Idea: add m_{H^+} as additional parameter to allow the network to account for changing kinematics across range of signal masses
- Benefits: more data & no need to train a network for each mass point
- Improve performance (overall, and on 'unseen' masses)
- But unclear what 'mass' to give to data samples in final fit, & might require mass-hypothesis dependent background estimation
 - → **not included in final analysis**



(a)



(b)

Figure 5: RoC AuC scores showing the performance of example $qqbb$ (a) and $lvbb$ (b) single-mass vs. parametrized vs. non-parametrized-all-mass NNs

Future consideration: Transformer-like NN for classification

- Idea: including lower-level information could allow the network to do better than we can with 'clever physics variables'
- Transformers are a natural choice for this:
 - ① Designed to learn relationships between input features
 - ② Can easily/naturally handle different numbers/types of object
- Encode 4-momenta & tag information, along with type-embedding ($e/\mu/\nu$ /small-R jet/large-R jets)
- Multiple layers of 'self-attention' (& optional MLP layers) followed by pooling, final classifier layers decide event as $lvbb$, $qqbb$ or background
- See big boost in performance
 - **2x** efficiency in SRs (for same background) at high-mass, **10x** at low-mass

Future consideration: Transformer-like NN for reconstruction

- Motivation: current reconstruction limited: $<60\%$ reco-efficiency, 20% misclass-rate
 - Idea: use similar architecture as previous slide (without pooling) to **reconstruct** the $H^+ \rightarrow W^+ h$ decay
 - Each object classified as belonging to SM-Higgs-from- H^+ , W^+ -from- H^+ , or neither
-
- See boost in reco-efficiency to $\sim 90\%$ as well as large reduction in mis-reconstruction

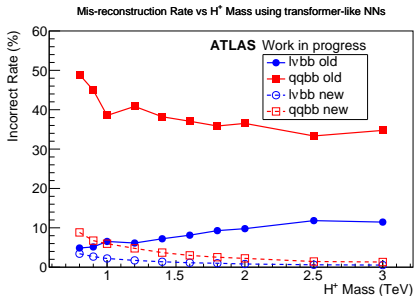


Figure 6: Percentage of signal reconstructed into incorrect class. Across all mass points, average of 20% is lost

Future consideration: transformer-like networks results

- 90% 'perfect', truth-matched signal-reco efficiency
- Leads to more background reconstructed, but ...
- Combined with classifier, improve S/\sqrt{B} by factors of 1.6 in most high-purity signal regions (up to 10 in some low-mass regions)
- Improvement in stats for signal regions & better mass reconstruction \rightarrow increased significance / decreased expected exclusion limits (see fig. 7)

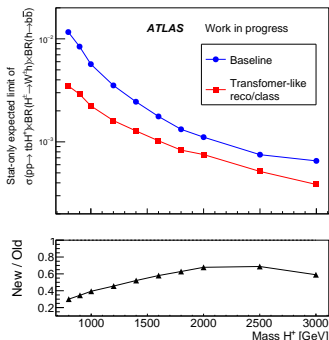


Figure 7: Stat only asimov fit expected 95% confidence limits of the transformer-like approach compared to baseline analysis strategy

Back to the present: Statistical fit

		Merged	
High-NN score signal region	<i>b</i> -tag multiplicity Mass window NN score	$w_{NN}^{lbbb} \geq 0.83$	$0b, \geq 1b$ $95 \text{ GeV} \leq m_J < 140 \text{ GeV}$ $w_{NN}^{qqbb} \geq 0.2$ (for events with $0b$) $w_{NN}^{qqbb} \geq 0.1$ (for events with $\geq 1b$)
Medium-NN score signal region	<i>b</i> -tag multiplicity Mass window NN score	$0.4 \leq w_{NN}^{lbbb} < 0.83$	$0b, \geq 1b$ $95 \text{ GeV} \leq m_J < 140 \text{ GeV}$ -
Low-NN score signal region	<i>b</i> -tag multiplicity Mass window NN score	$w_{NN}^{lbbb} < 0.4$	$0b, \geq 1b$ $95 \text{ GeV} \leq m_J < 140 \text{ GeV}$ $w_{NN}^{qqbb} < 0.2$ (for events with $0b$) $w_{NN}^{qqbb} < 0.1$ (for events with $\geq 1b$)
Low-mass control region	<i>b</i> -tag multiplicity Mass window NN score		$0b, \geq 1b$ $m_J < 95 \text{ GeV}$ -
High-mass control region	<i>b</i> -tag multiplicity Mass window NN score		$0b, \geq 1b$ $m_J \geq 140 \text{ GeV}$ -

- 18 regions
 - Sideband control regions with high/low m_h
 - Low, medium and high score Signal regions defined in terms of NN score ($qqbb$ skips medium region due to lack of stats)
- Fit m_{W+h}

Results

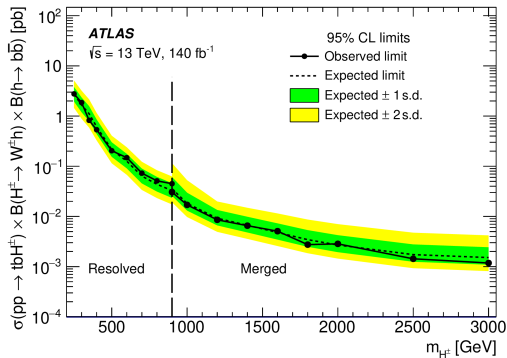


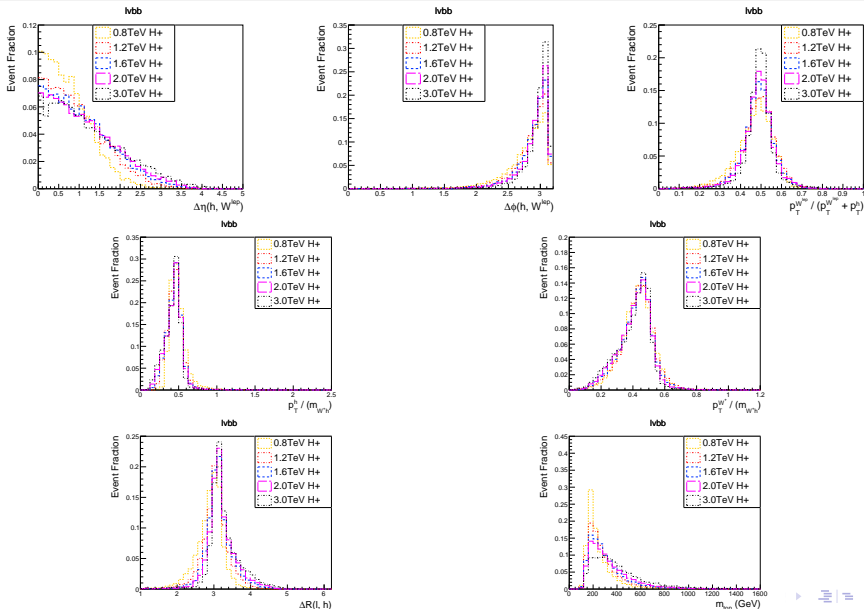
Figure 8: Combined limits presented alongside resolved channel. Boosted channel results become better at 900GeV and above

- No significant excess observed
- 95% Confidence interval limits set across the mass range

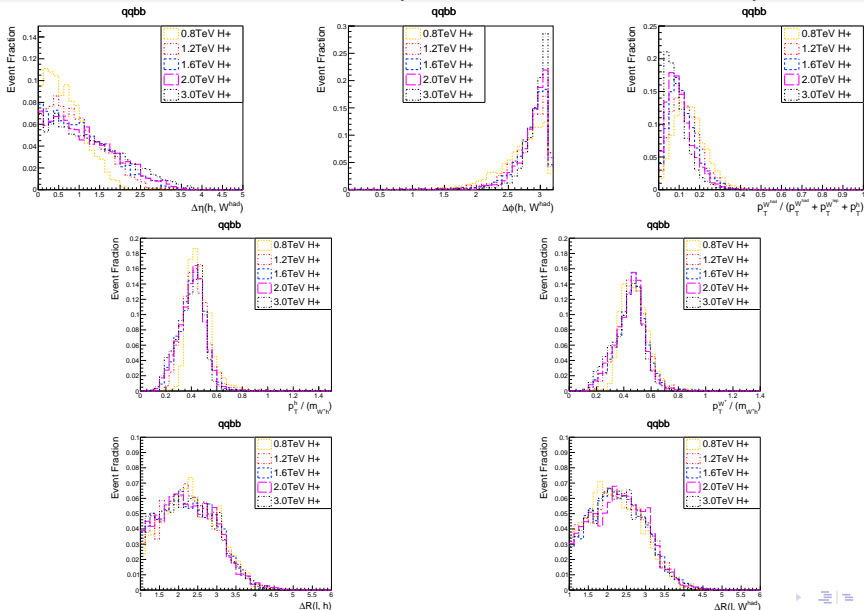
BACKUP

BACKUP

$lvbb$ NN inputs (multiple signal masses)



$q\bar{q}b\bar{b}$ NN inputs (multiple signal masses)



Baseline NN outputs ($m_{H^+} = 0.8$ TeV)

Output score histograms for $l\nu bb$ (figs. 9a and 9b) and $qqbb$ (figs. 9c and 9d) NNs (using the baseline method)

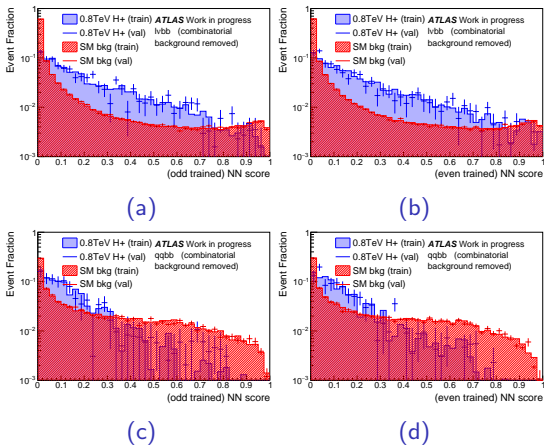


Figure 9: Distributions of background (red) and signal (blue) samples for the $l\nu bb$ and $qqbb$ NNs

Baseline NN outputs ($m_{H^+} = 1.2$ TeV)

Output score histograms for $l\nu bb$ (figs. 10a and 10b) and $qqbb$ (figs. 10c and 10d) NNs (using the baseline method)

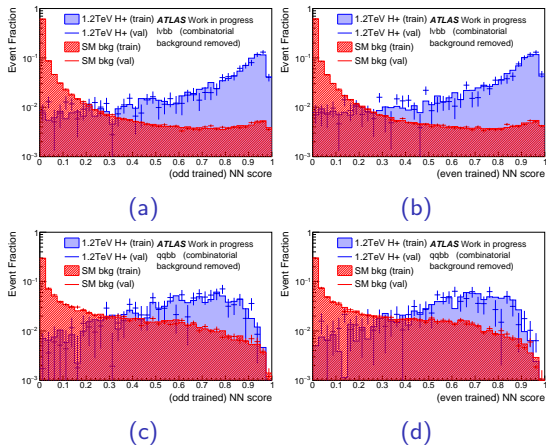


Figure 10: Distributions of background (red) and signal (blue) samples for the $l\nu bb$ and $qqbb$ NNs

Baseline NN outputs ($m_{H^+} = 3.0$ TeV)

Output score histograms for $l\nu bb$ (figs. 11a and 11b) and $qqbb$ (figs. 11c and 11d) NNs (using the baseline method)

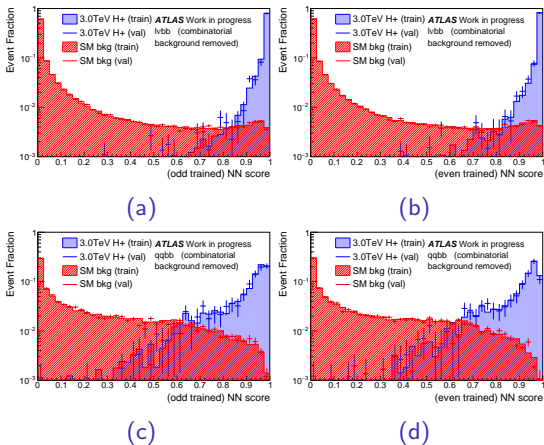
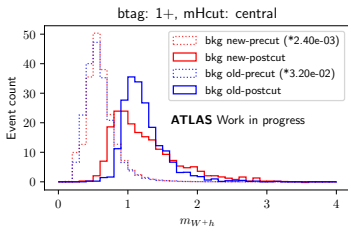
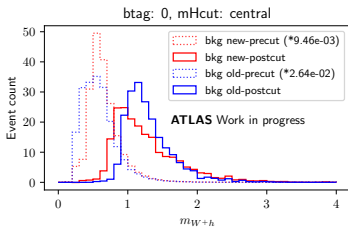


Figure 11: Distributions of background (red) and signal (blue) samples for the $l\nu bb$ and $qqbb$ NNs.

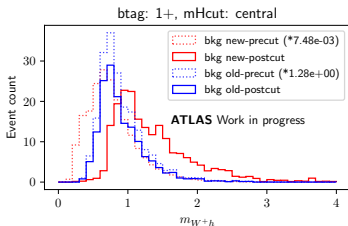
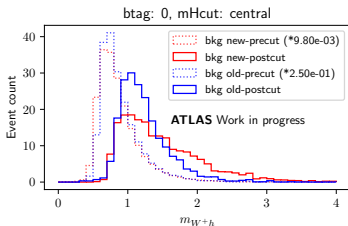
Transformer mass sculpting

Plots of background remaining before/ after cut to retain 200 bkg samples

lvbb



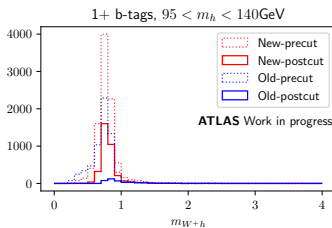
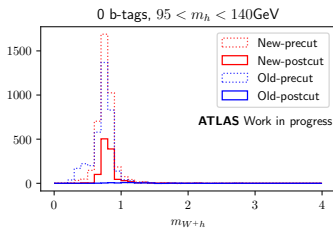
qqbb



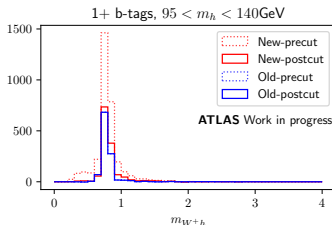
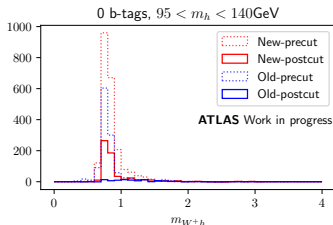
Transformer performance ($m_H^+ = 0.8 \text{ TeV}$)

Plots of signal remaining before/ after cut to retain 200 bkg samples

lvbb 0.8 TeV H+



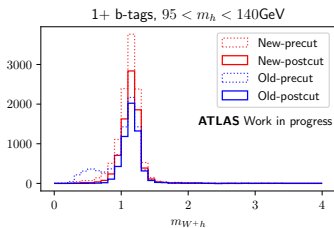
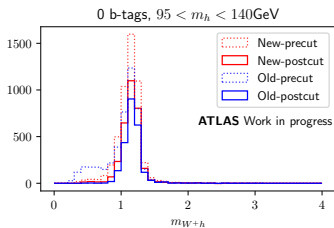
qqbb 0.8 TeV H+



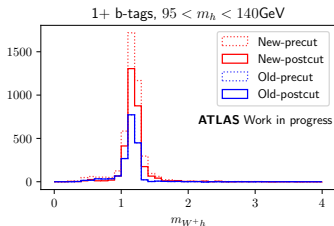
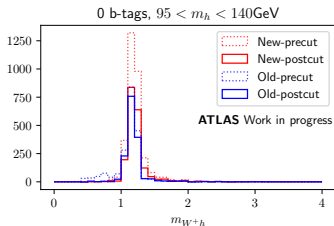
Transformer performance ($m_H^+ = 1.2 \text{ TeV}$)

Plots of signal remaining before/ after cut to retain 200 bkg samples

lvbb 1.2 TeV H+



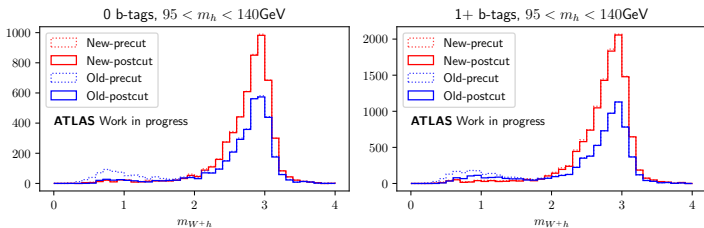
qqbb 1.2 TeV H+



Transformer performance ($m_H^+ = 3.0$ TeV)

Plots of signal remaining before/ after cut to retain 200 bkg samples

lvbb 3.0 TeV H+



qqbb 3.0 TeV H+

