Taggers for boosted HH searches within the ATLAS experiment

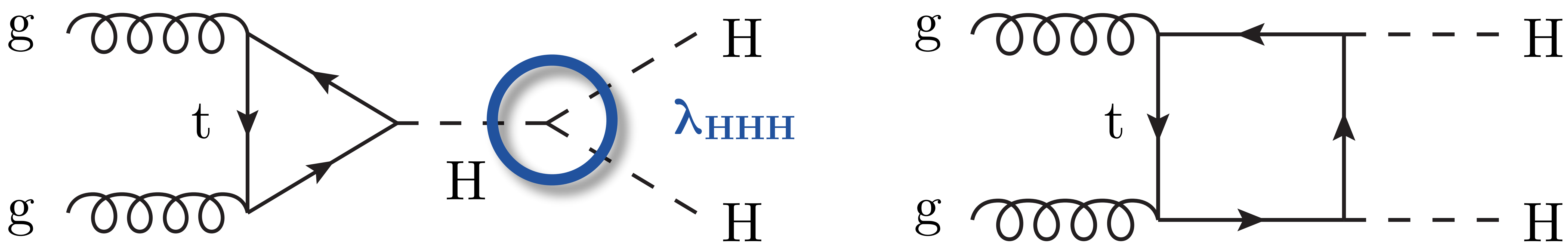
Luca Cadamuro

IJCLab, CNRS/IN2P3, Université de Paris-Saclay

February 28th, 2024

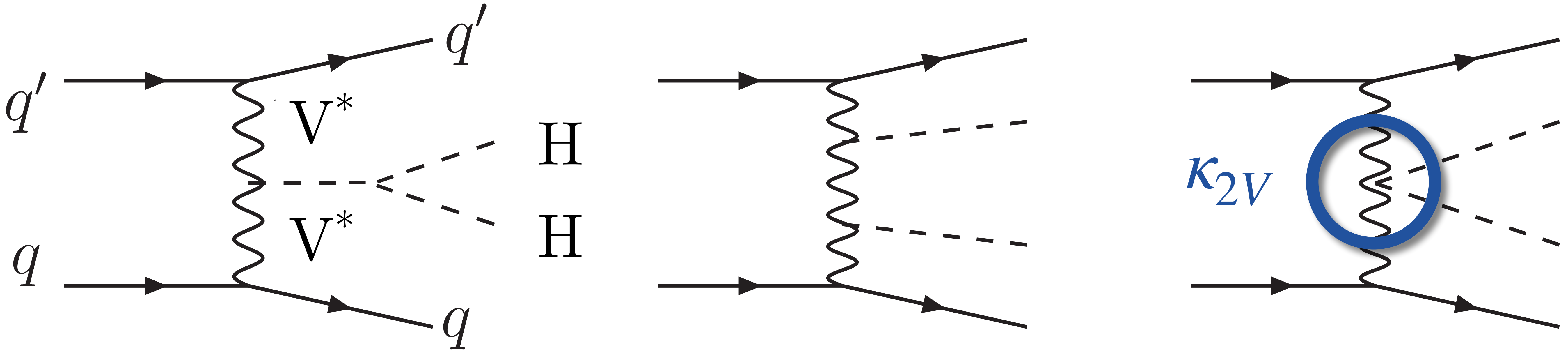
1st COMETA general meeting - İzmir

HH physics motivations



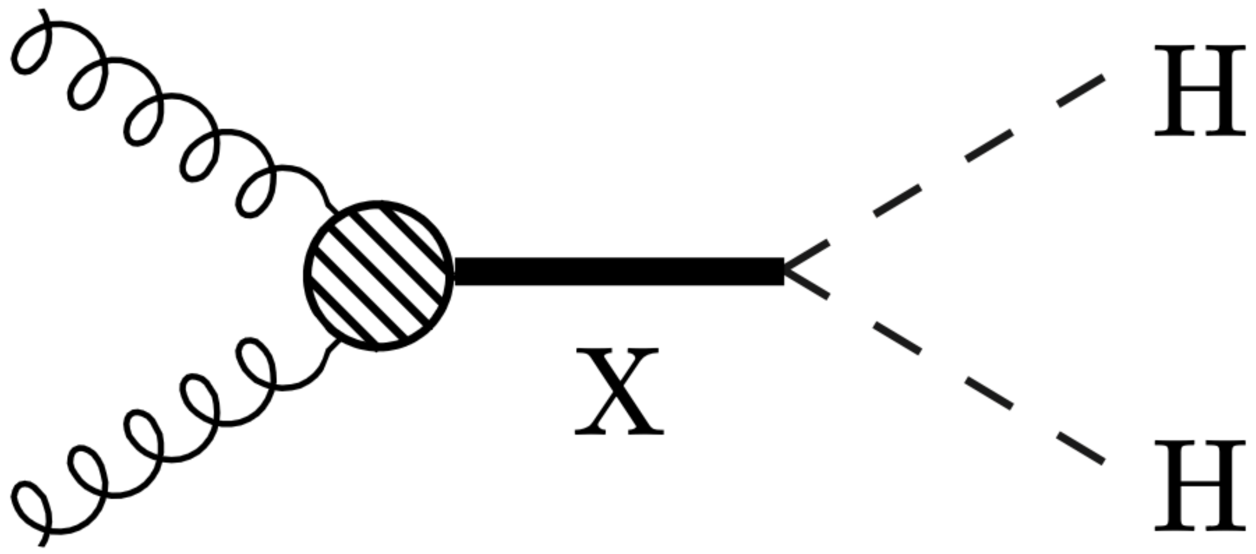
Gluon fusion

SM HH signal
Self-coupling (κ_λ), EFT



Vector boson fusion

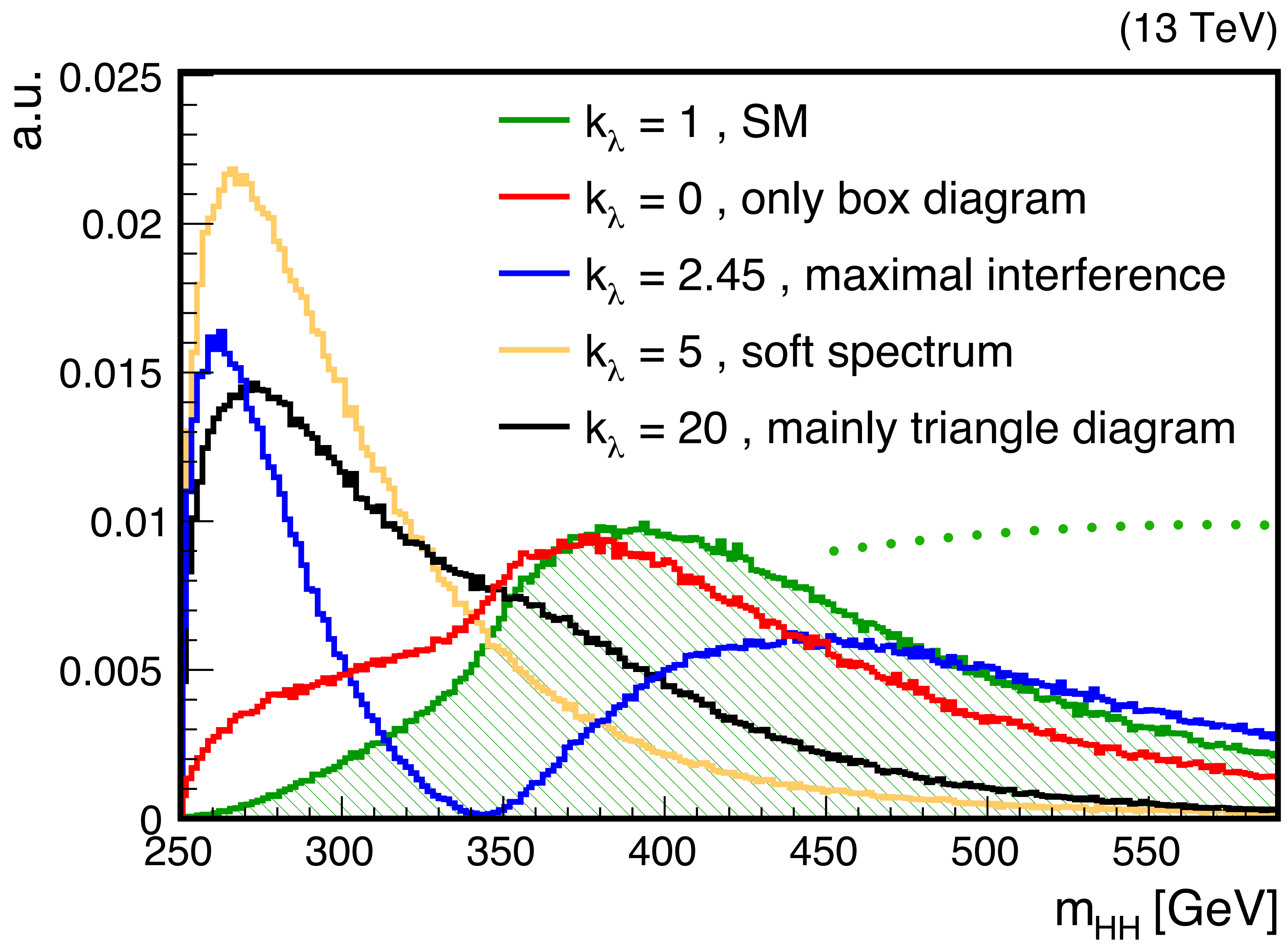
VVHH coupling (κ_{2V})

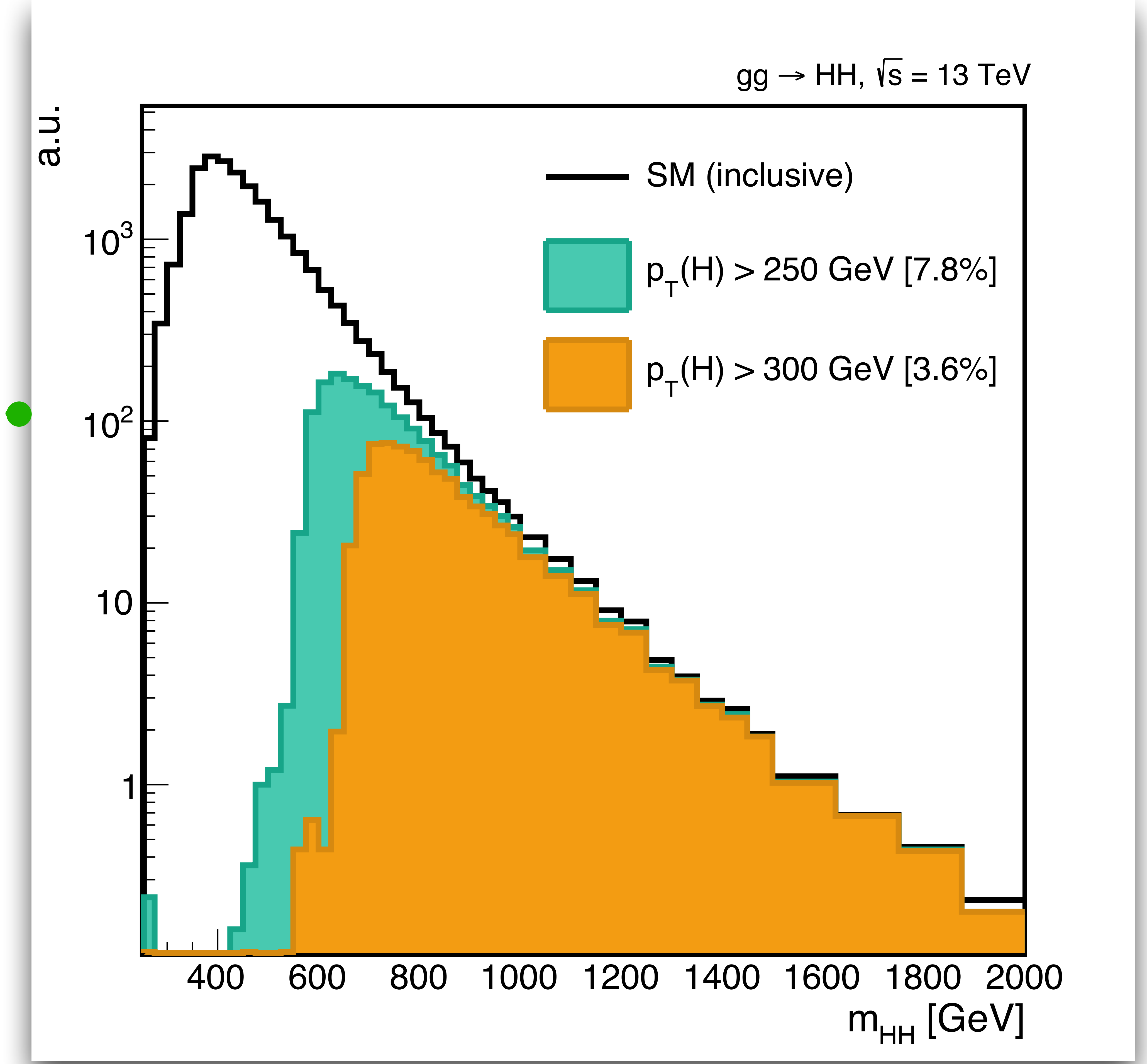


Resonant production

New heavy states

Boosted HH in ggF production





- SM has one of the hardest m_{HH} spectra (κ_λ effects typically at threshold)

Tiny phase space portion that can provide advantageous S/B separation

Gluon fusion

SM HH signal
Self-coupling (κ_λ), EFT

Vector boson fusion

VVHH coupling (κ_{2V})

Resonant production

New heavy states

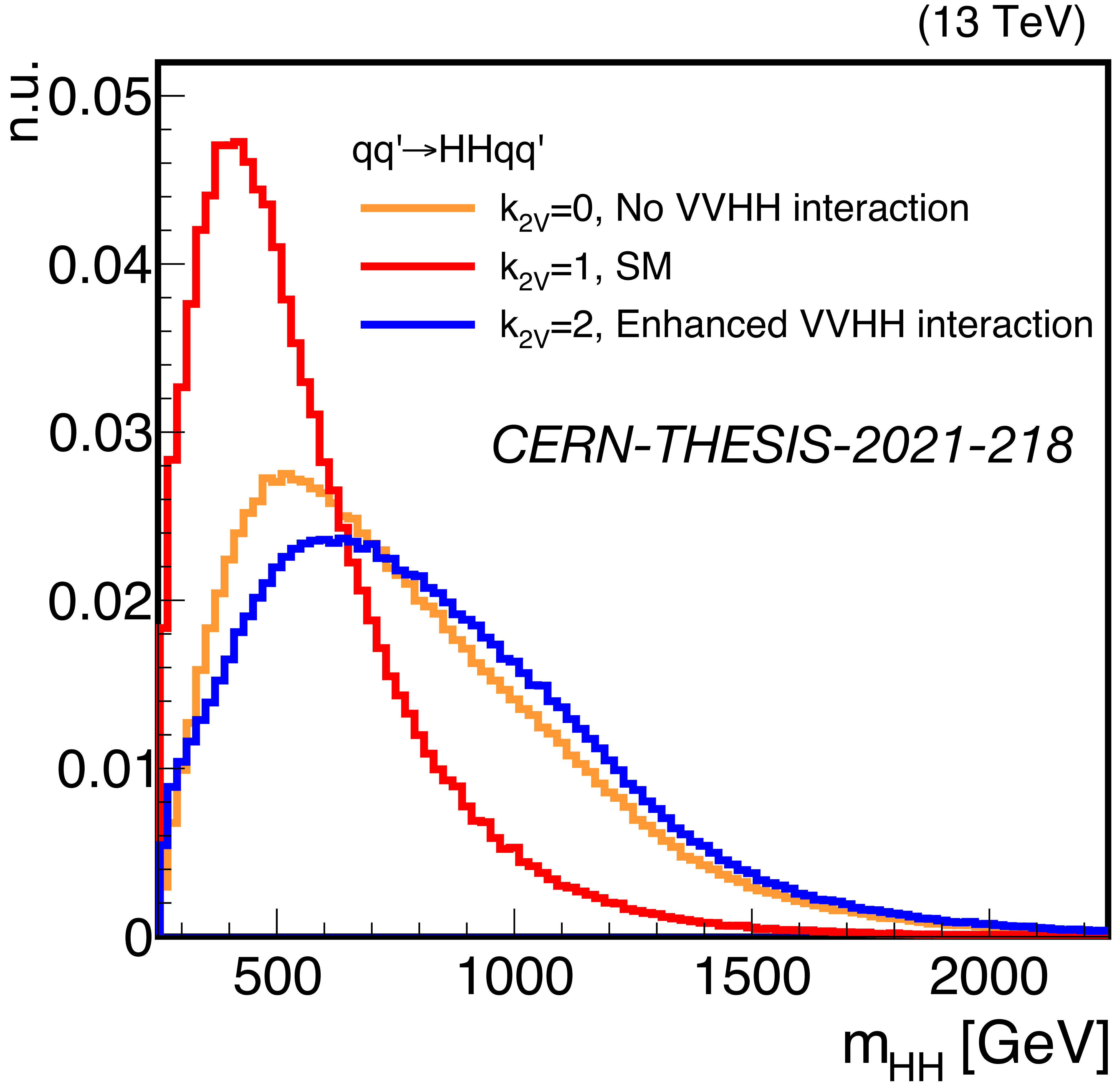
Boosted HH in VBF production

$$\mathscr{A}\left(V_L V_L \to \mathrm{HH}\right) \simeq \frac{\hat{s}}{v^2}\left(\kappa_{2V} - \kappa_V^2\right)$$

EPJC 77 (2017) 7, 481

- Very large changes in m_{HH} with O(1) κ_{2V} variations because of alterations in the cancellations from electroweak doublet structure

Dominant boosted production in case of anomalous couplings



Gluon fusion

SM HH signal
Self-coupling (κ_λ), EFT

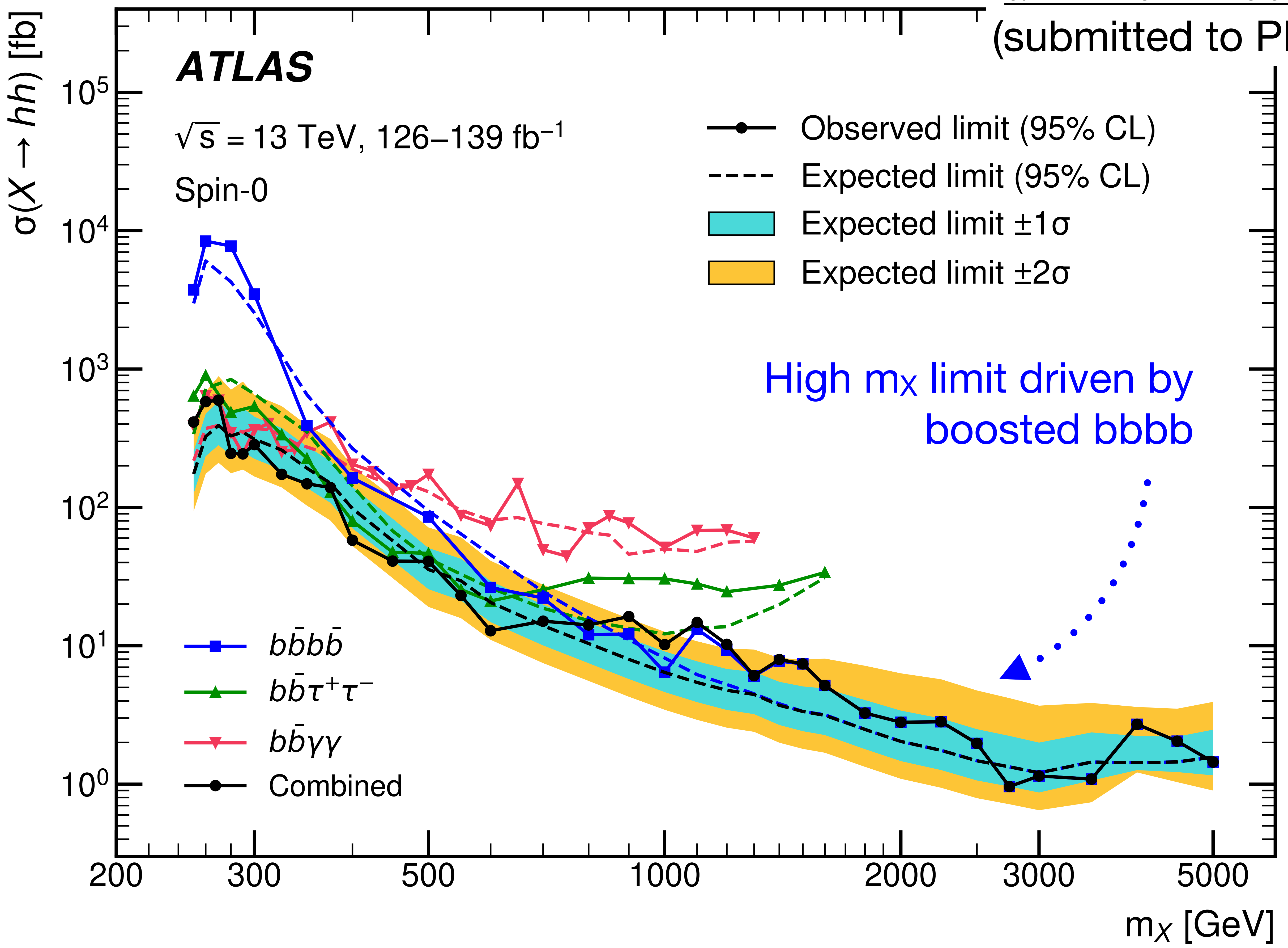
Vector boson fusion

VVHH coupling (κ_{2V})

Resonant production

New heavy states

Boosted HH in resonant production



arXiv:2311.15956 (submitted to PRL)

- Resonant searches span a broad m_X range
- High mass resonances result in highly boosted H bosons
- Boosted H tagging is fundamental to explore $m_X > 1$ TeV

Boosted topologies naturally dominant at high m_X

Gluon fusion

SM HH signal
Self-coupling (κ_λ), EFT

Vector boson fusion

VVHH coupling (κ_{2V})

Resonant production

New heavy states

Which final state for boosted HH?

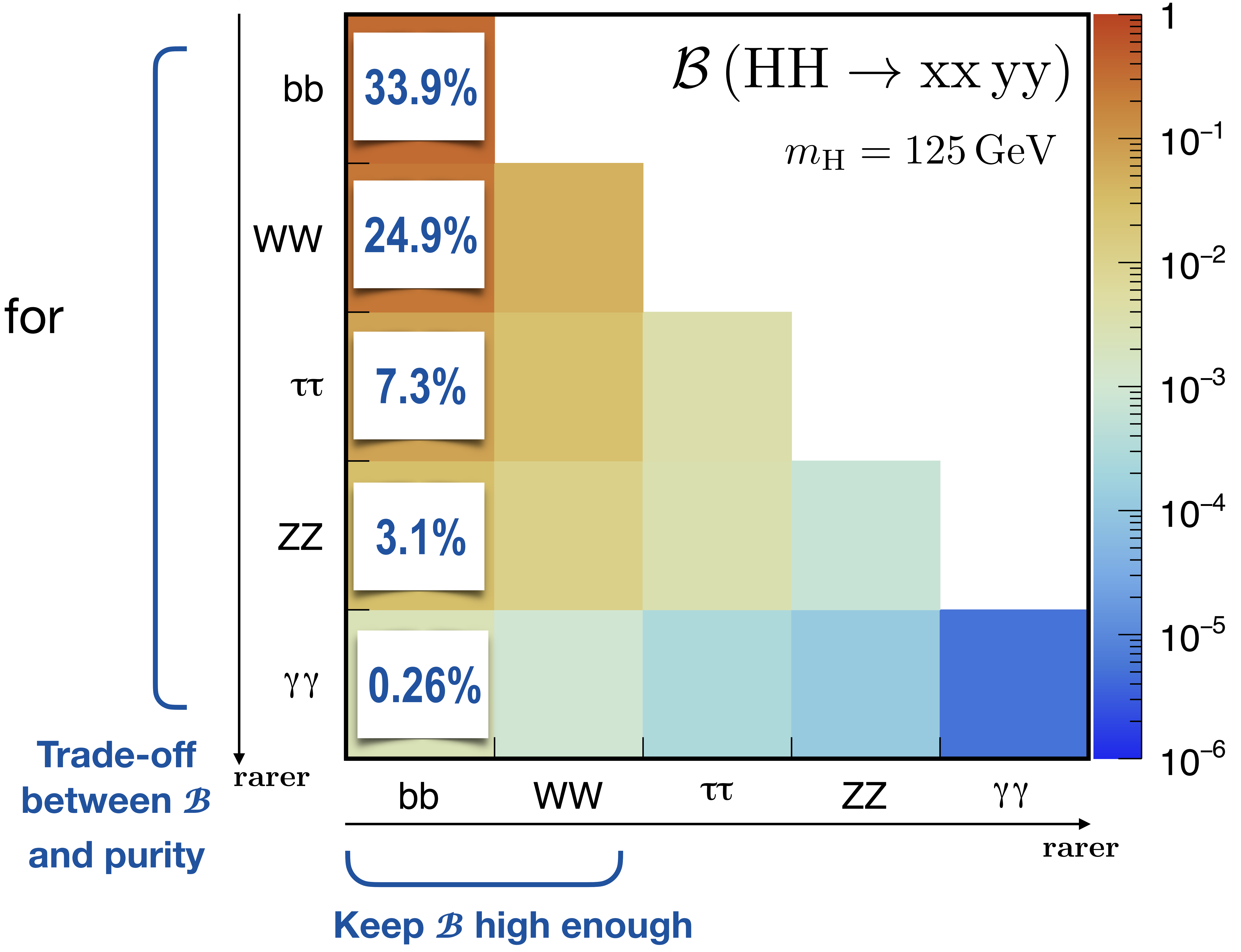
- HH searches span a broad set of final states
- Boosted topologies usually identify a small portion of the phase space (ggF) or feature small cross sections (VBF, resonant)
- High BR channels with hadrons most suited for the application of boosted tagging in HH

H→ bb/WW/$\tau\tau$ final states are particularly interesting for boosted HH searches

Boosted W final states were not yet explored in an ATLAS HH search and hence boosted W tagging is not covered here.

Several boosted W algorithms exist and could be applied for this topology.

XX % : ATLAS public results with full Run 2 data set



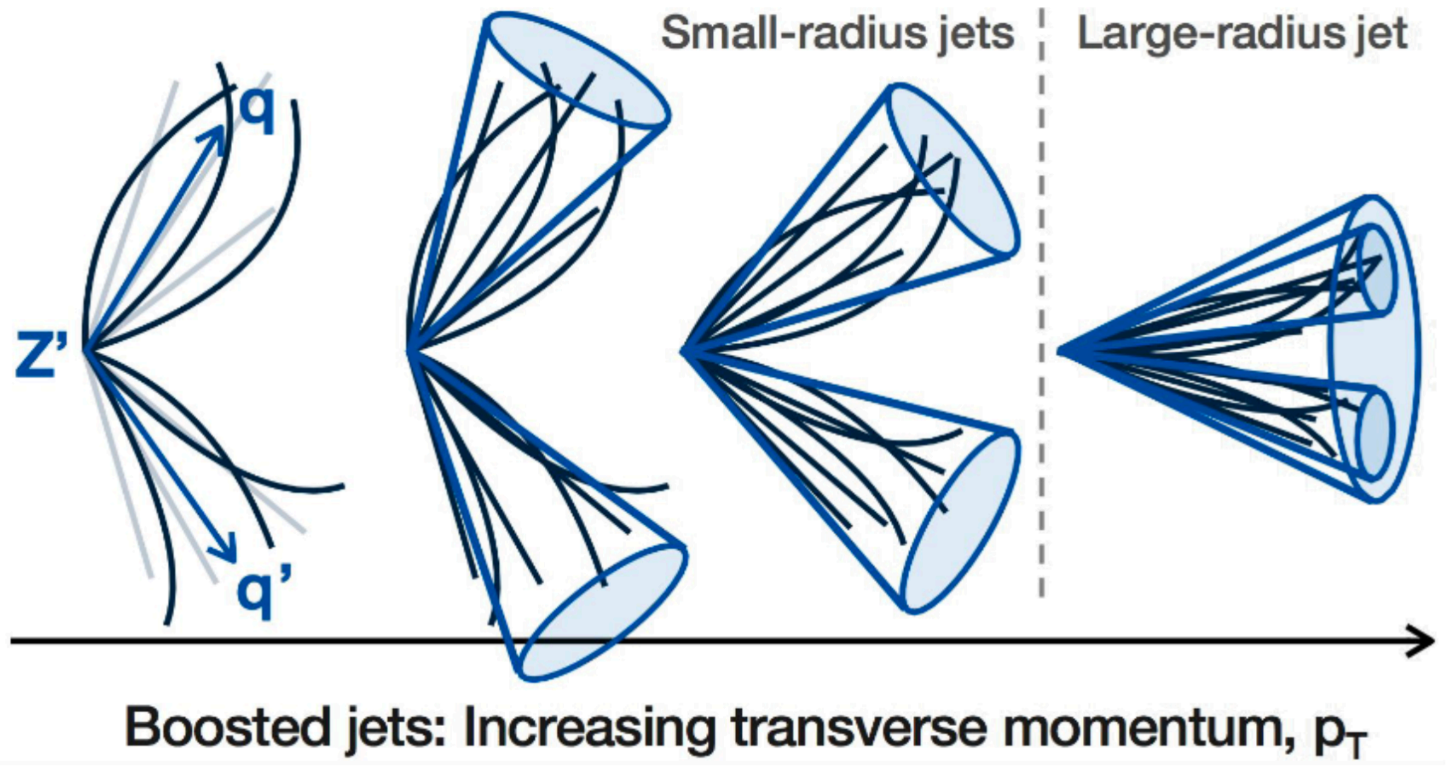
Boosted tagging inputs : UFO jets

- Unified Flow Objects jets combine optimally calorimeter and tracker information (EPJC 81 (2021) 4, 334)
  - Particle Flow Objects (PFO) : particle energy estimation from track subtracted from calo cluster
  - Track-Calo Cluster (TCC) : energy from calo cluster + η,ϕ information from tracks
  - dedicated pileup mitigation and jet grooming

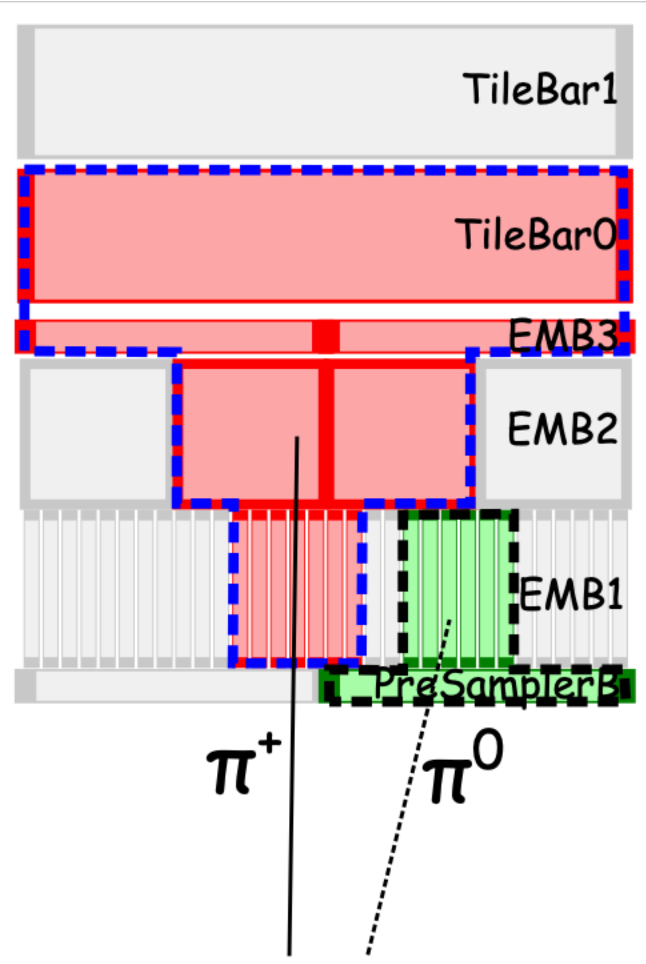
Clustering done with R = 1

$\Delta R \sim 2m/p_T \rightarrow$ for Higgs, boosted reconstruction from $p_T(H) \approx 250$ GeV

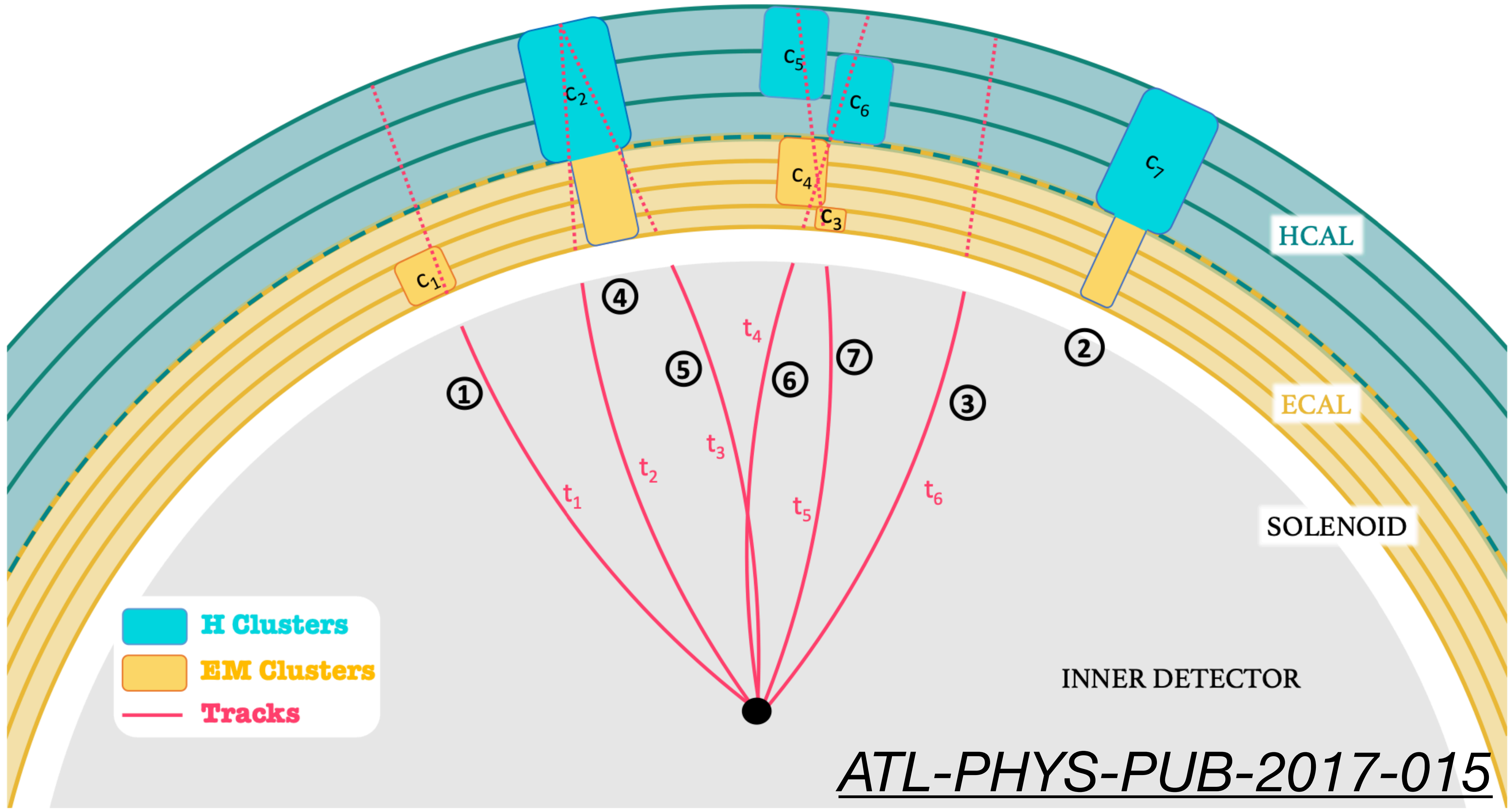
Subjets identified by clustering constituents as variable radius (VR) jets with $R = \rho/p_T$



Input constituents as close as possible to individual physics particles



EPJC 77 (2017) 466



ATL-PHYS-PUB-2017-015

Outline of boosted bb taggers

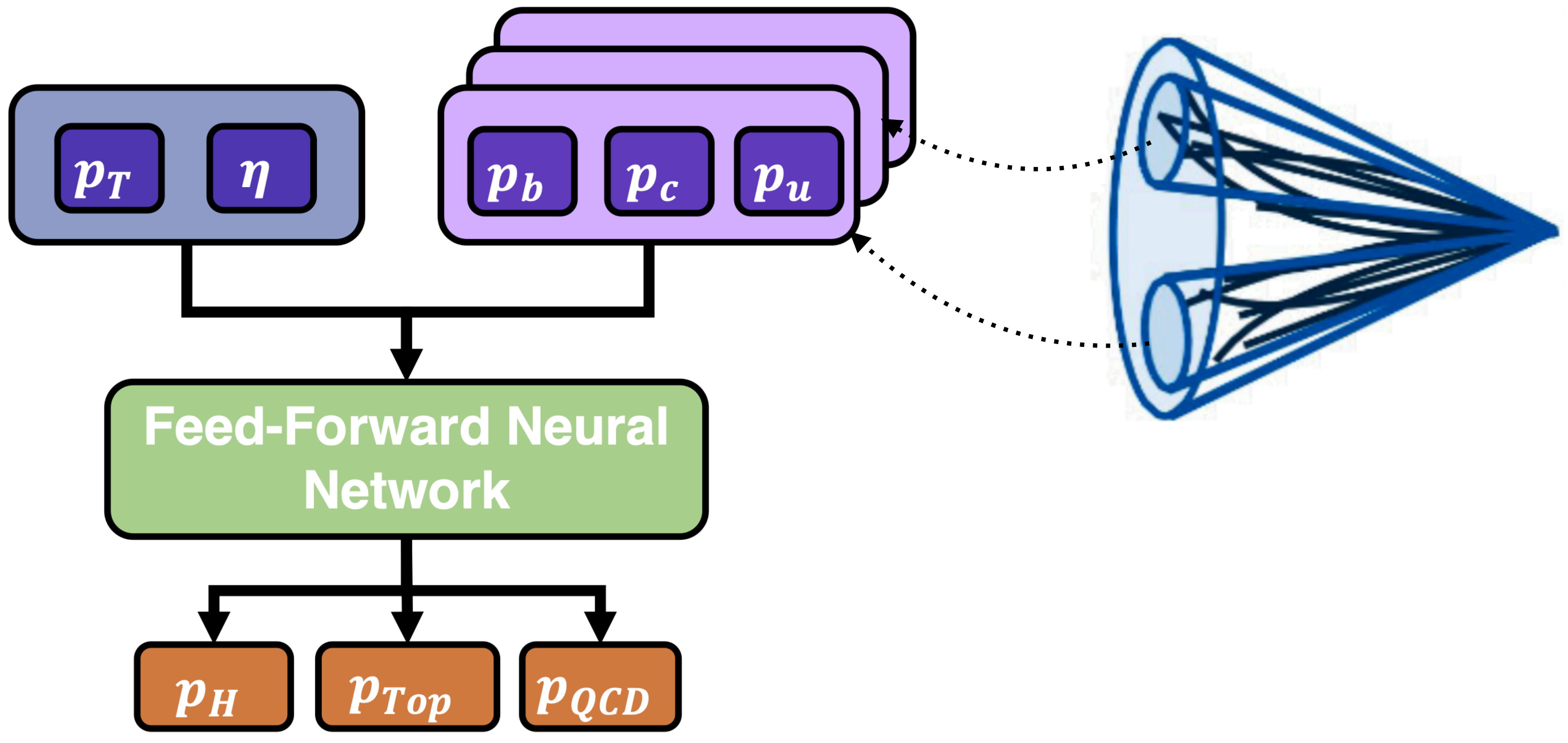
Taggers improve quickly following the technical evolutions in the field

Two main taggers are described in the following

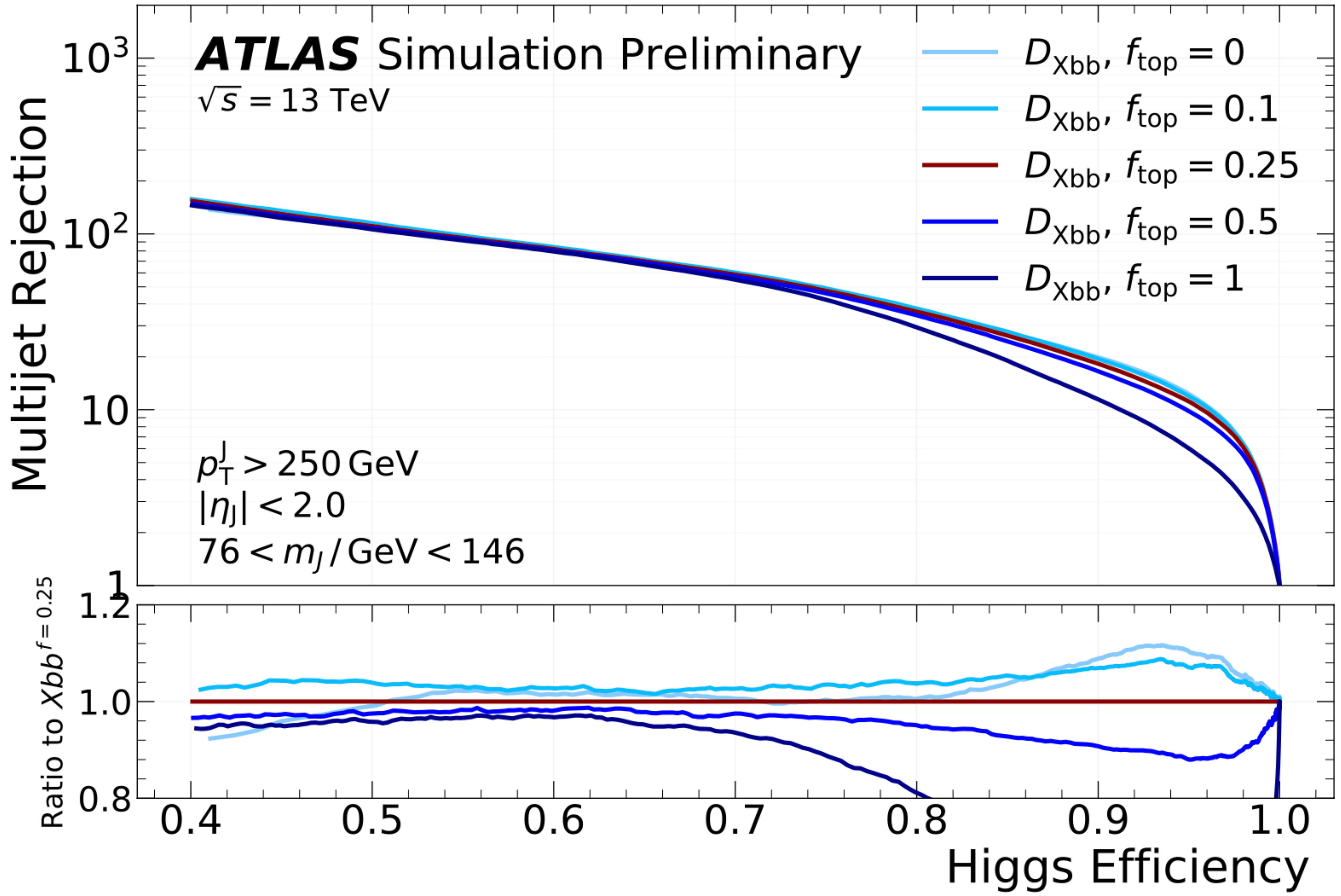
- Xbb tagger
  - feed-forward NN that combined flavour tagging discriminants from subjets
  - calibrated on ATLAS Run 2 data
  - used in most of the Run 2 results
- GN2X tagger
  - most recent development
  - uses low-level information from jet constituents in a GNN / Transformer network architecture

The comparison of their performance illustrates the power of going towards constituent-based taggers

Xbb tagger



$$D_{\text{Xbb}} = \ln \frac{p_{\text{Higgs}}}{f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{\text{top}}) \cdot p_{\text{multijet}}}$$

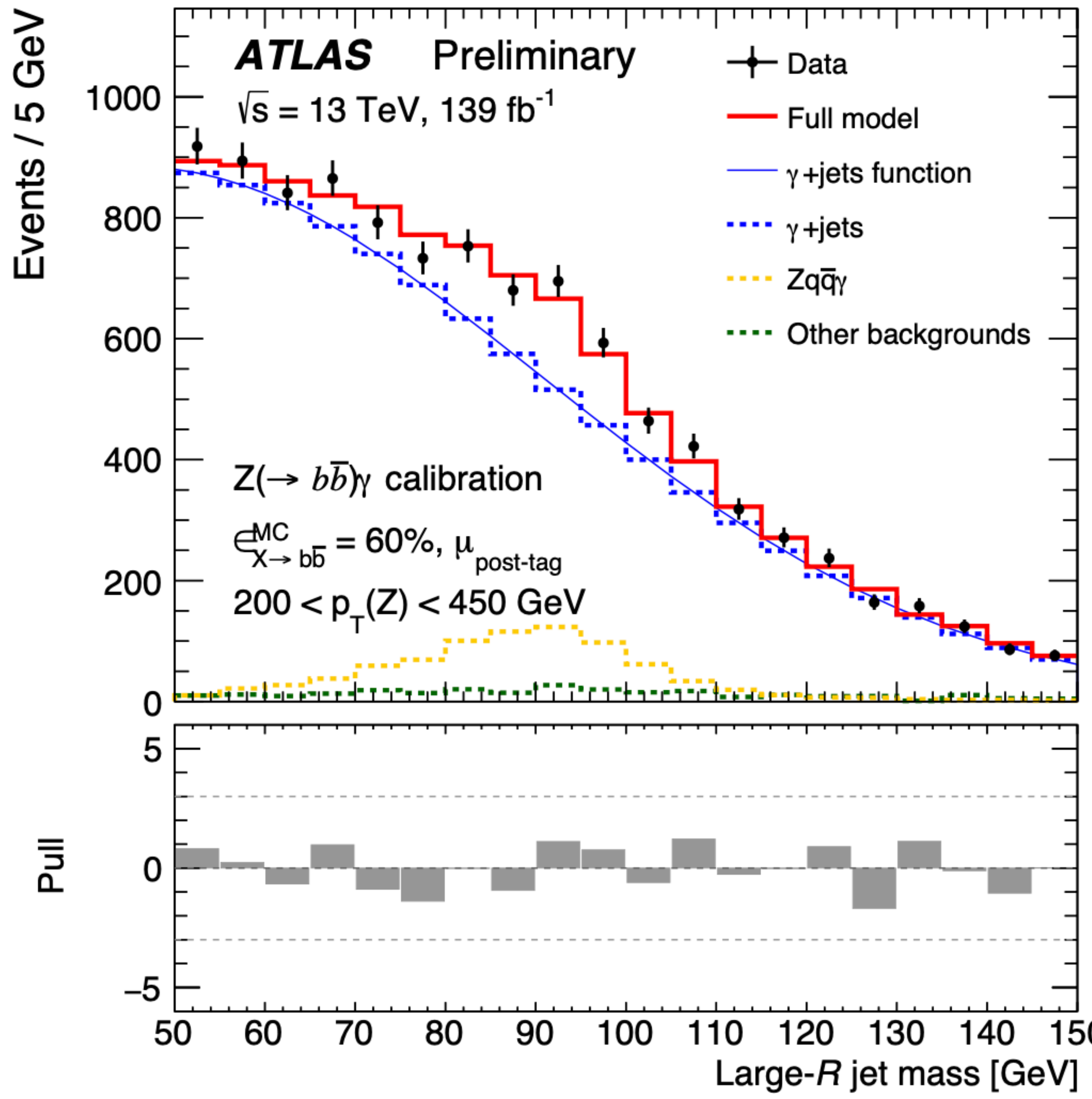


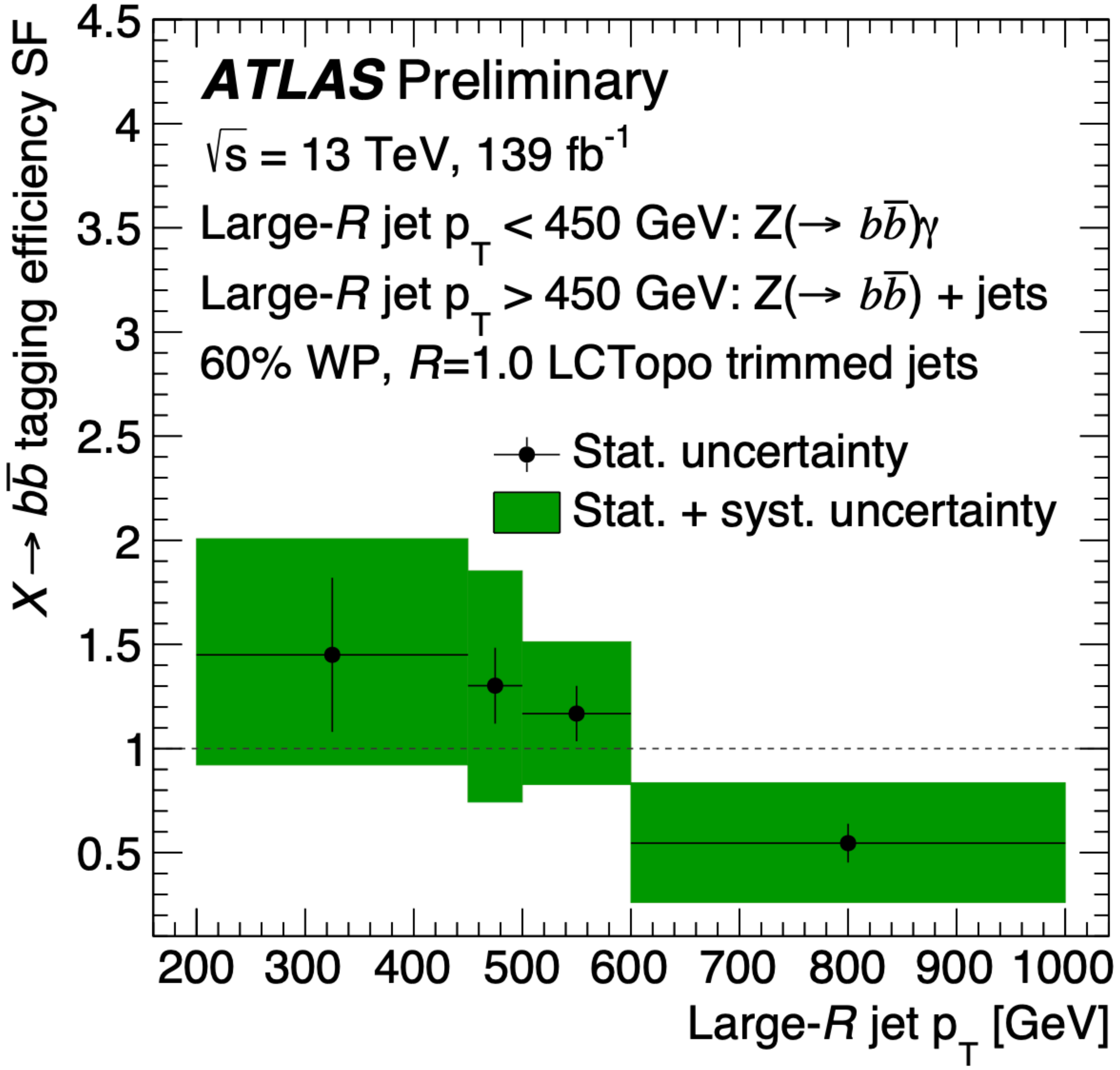
- Individual subjets are tagged using single b-tagging DL1r algorithm (DNN) optimised for VR jets
- Jet information + DL1r output nodes of max 3 subjets are fed to Xbb

Xbb tagger calibration

ATL-PHYS-PUB-2021-035

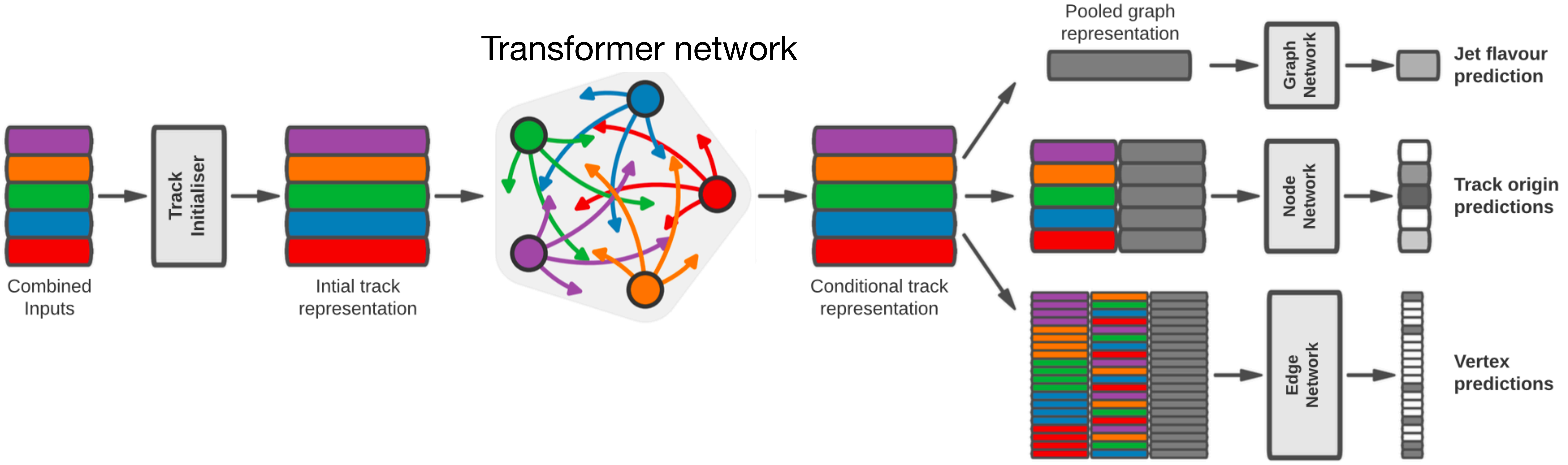
- Signal calibration on Z(bb)+jets and Z(bb)γ
- Background calibration on tt events
- Validation on g→bb events





GN2X

[ATL-PHYS-PUB-2022-027]



Large jet features : $p_T/\eta/m$

20 low-level track features (momentum, geometry, quality)

Embedding representation

Transformer Encoder

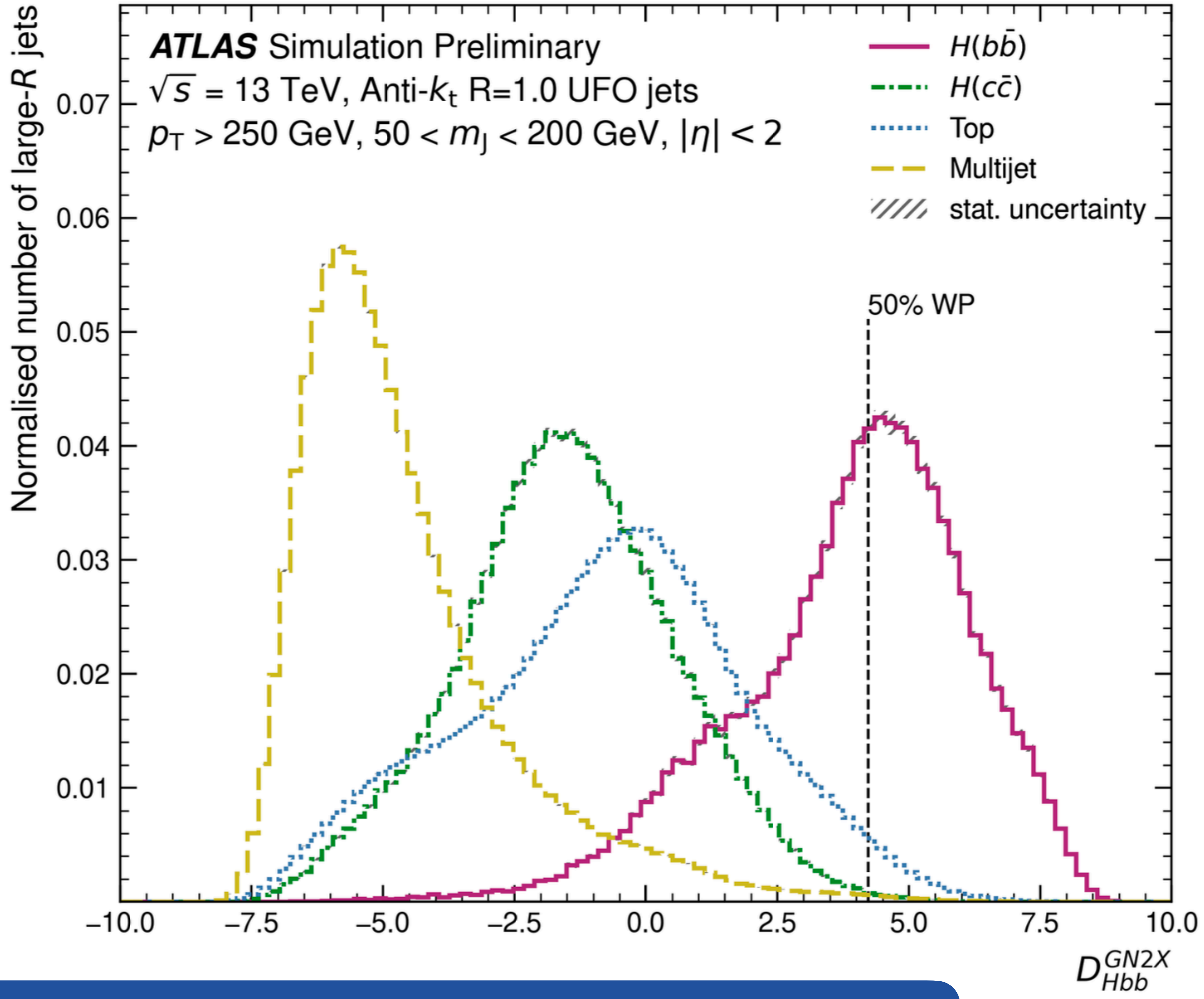
Primary task (jet flavour identification) + auxiliary tasks

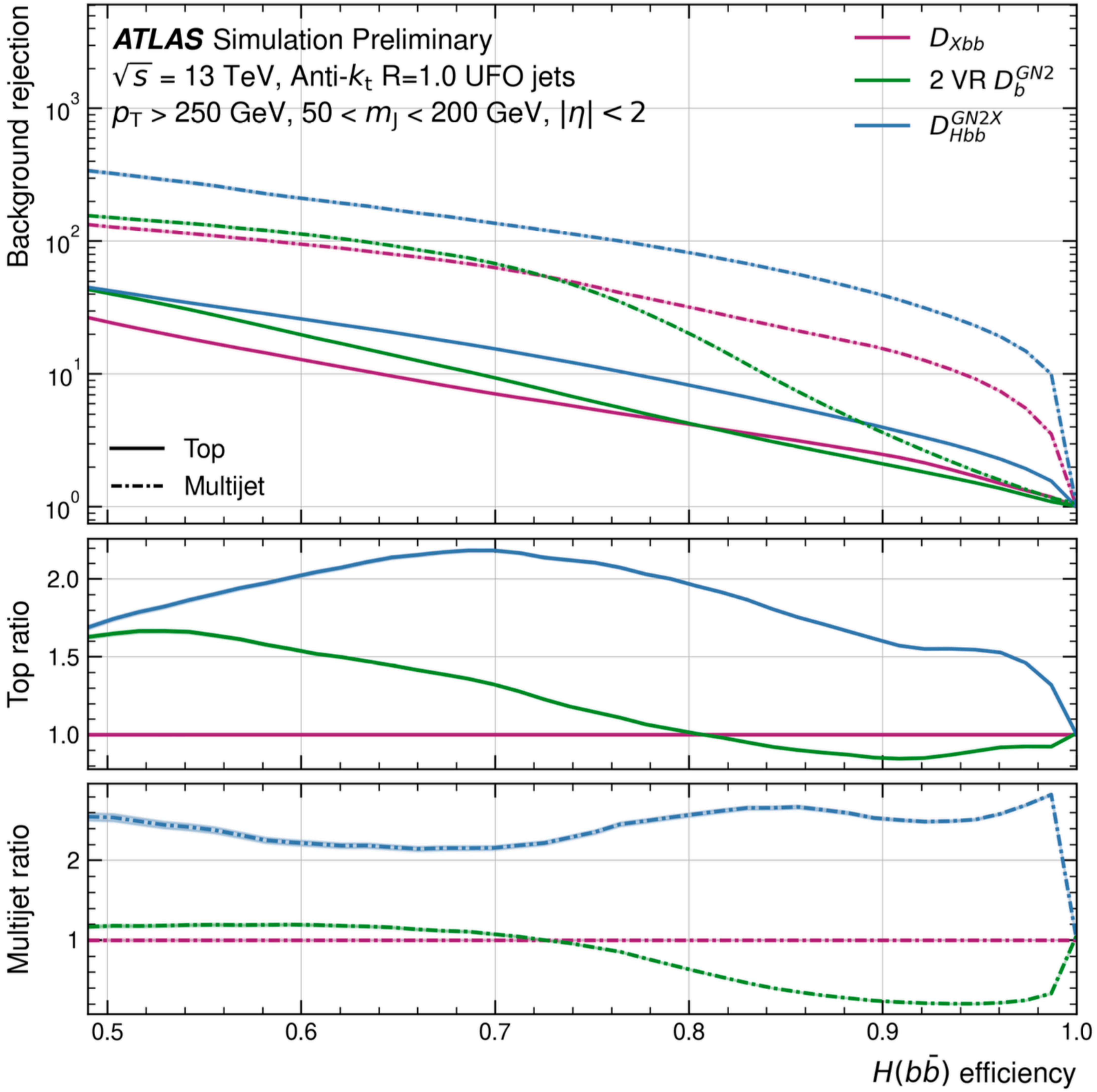
- Evolution of GN1 architecture (based on GNN)
- 4 target flavour classes (bb, cc, top, QCD)
- Training done on ~60M jets

Performance

$$D_{\mathrm{Hbb}}^{\mathrm{GN2X}} = \ln\left(\frac{p_{\mathrm{Hbb}}}{f_{\mathrm{Hcc}} \cdot p_{\mathrm{Hcc}} + f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{Hcc}} - f_{\mathrm{top}}) \cdot p_{\mathrm{QCD}}}\right)$$

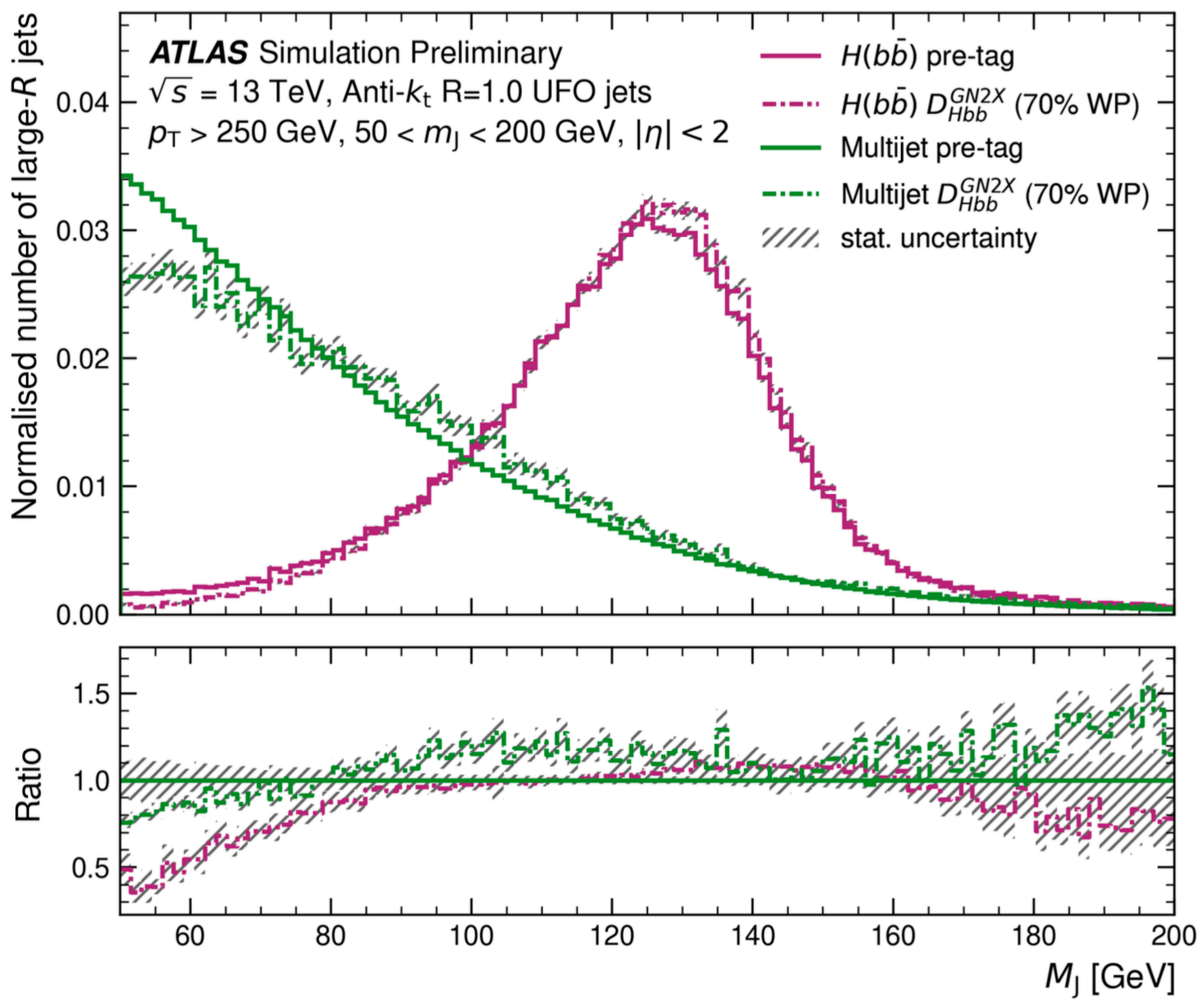
- GN2X vs Xbb: importance of low-level information
- GN2X vs 2 VR GN2 : importance of correlations across subjets





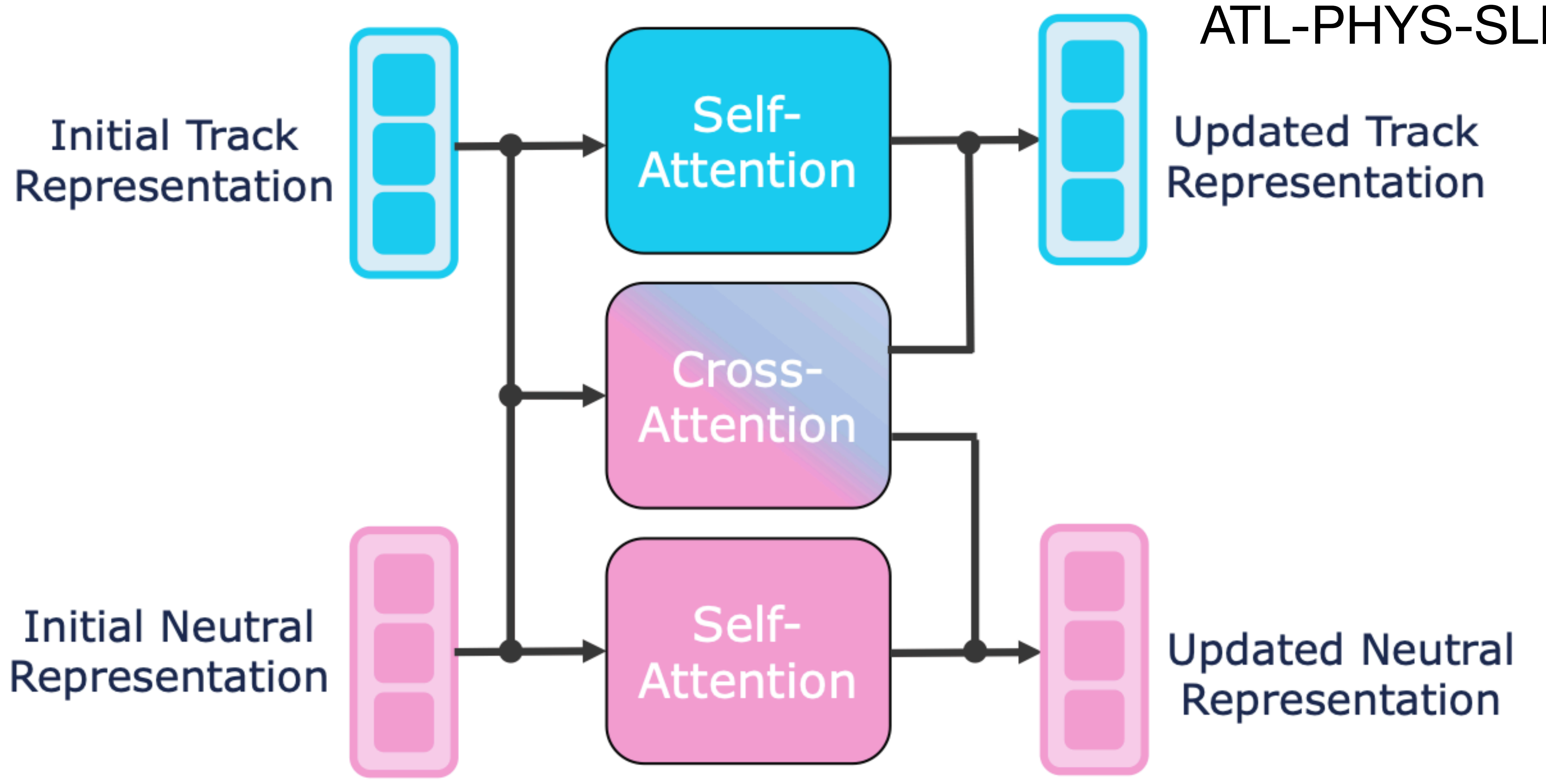
GN2X outperforms previous taggers

Mass sculpting

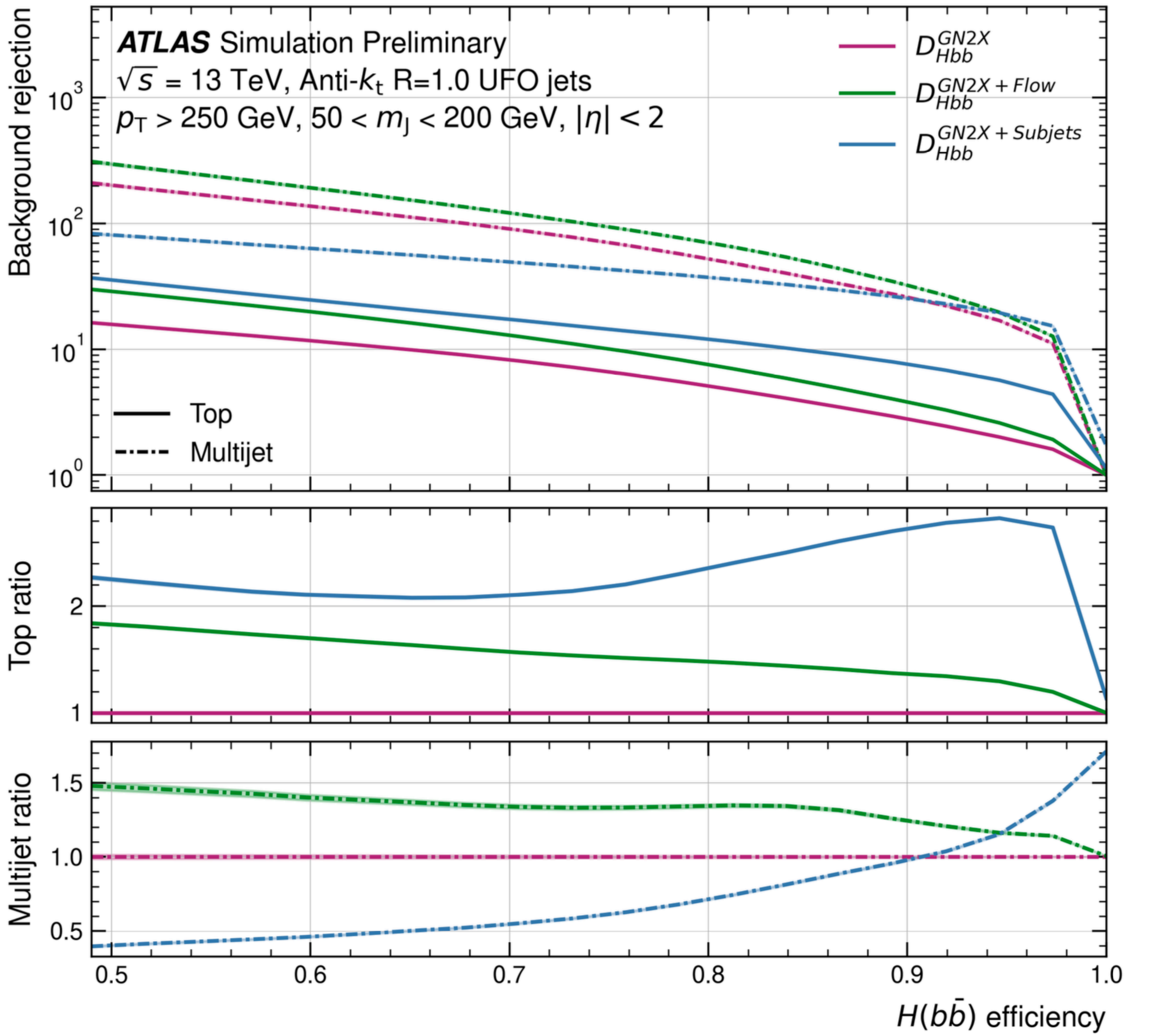


- Effect on the background observed in the edge mass regions
  - training performed with jets of mass [50, 300] GeV
  - approximately flattened signal sample from ZH(bb) and p_T-binned QCD sample

Beyond track information



ATL-PHYS-SLIDE-2023-328



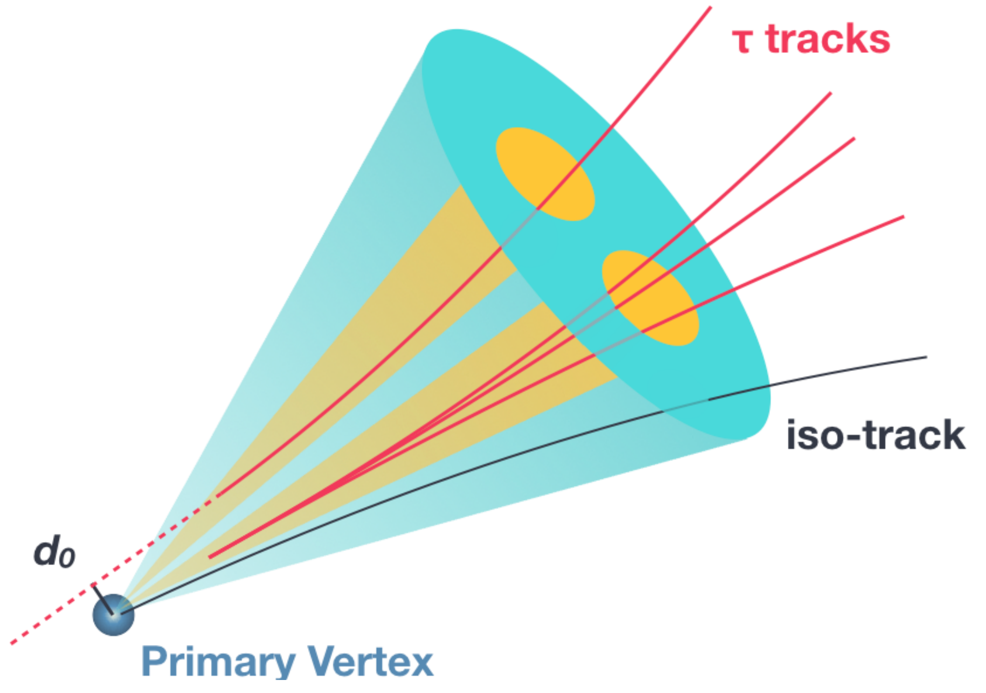
- Inclusion of neutral information (energy flow constituents) and subjet information further improves the performance

Boosted $H \to \tau\tau$ reconstruction

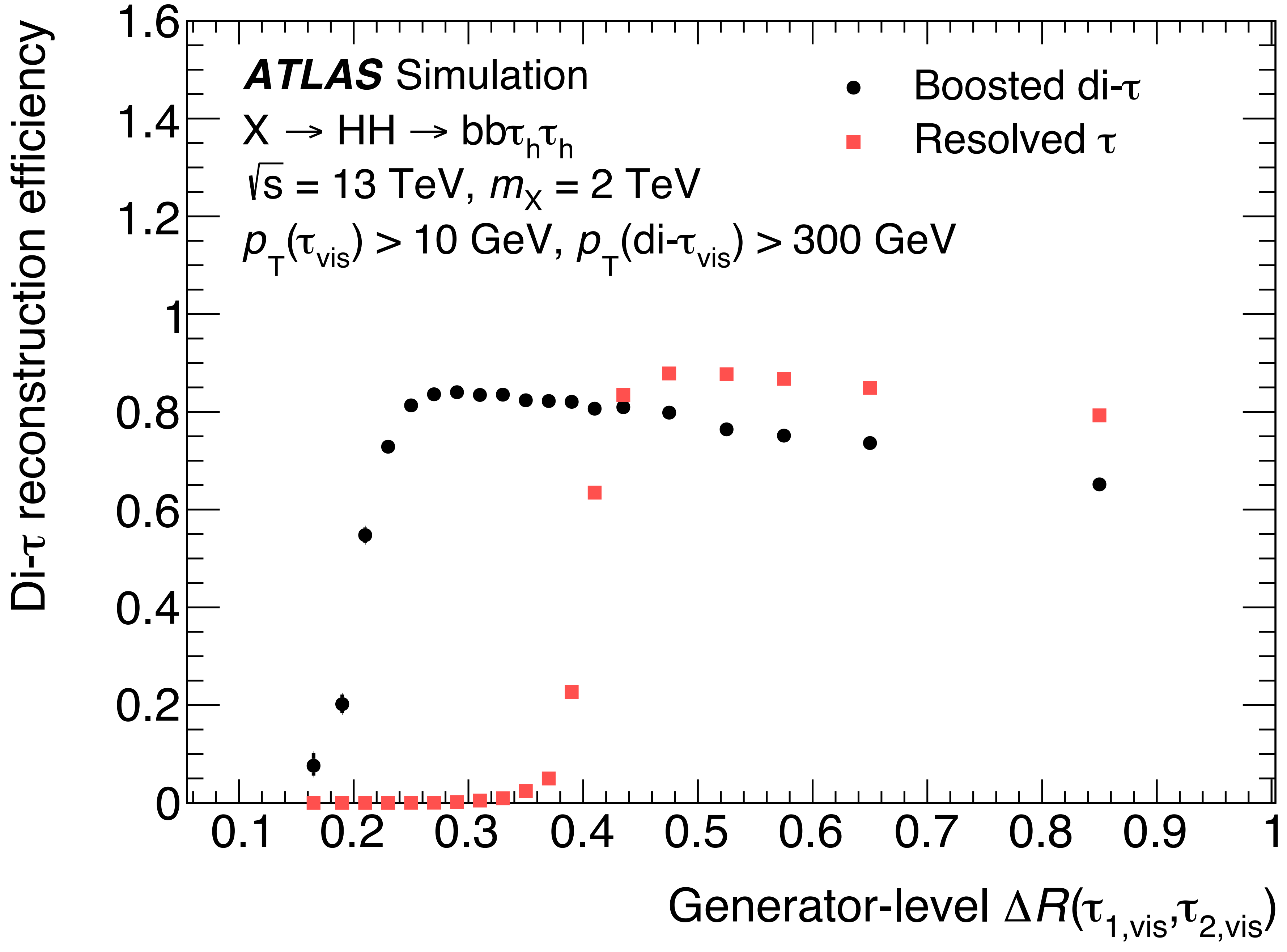
JHEP 11 (2020) 163

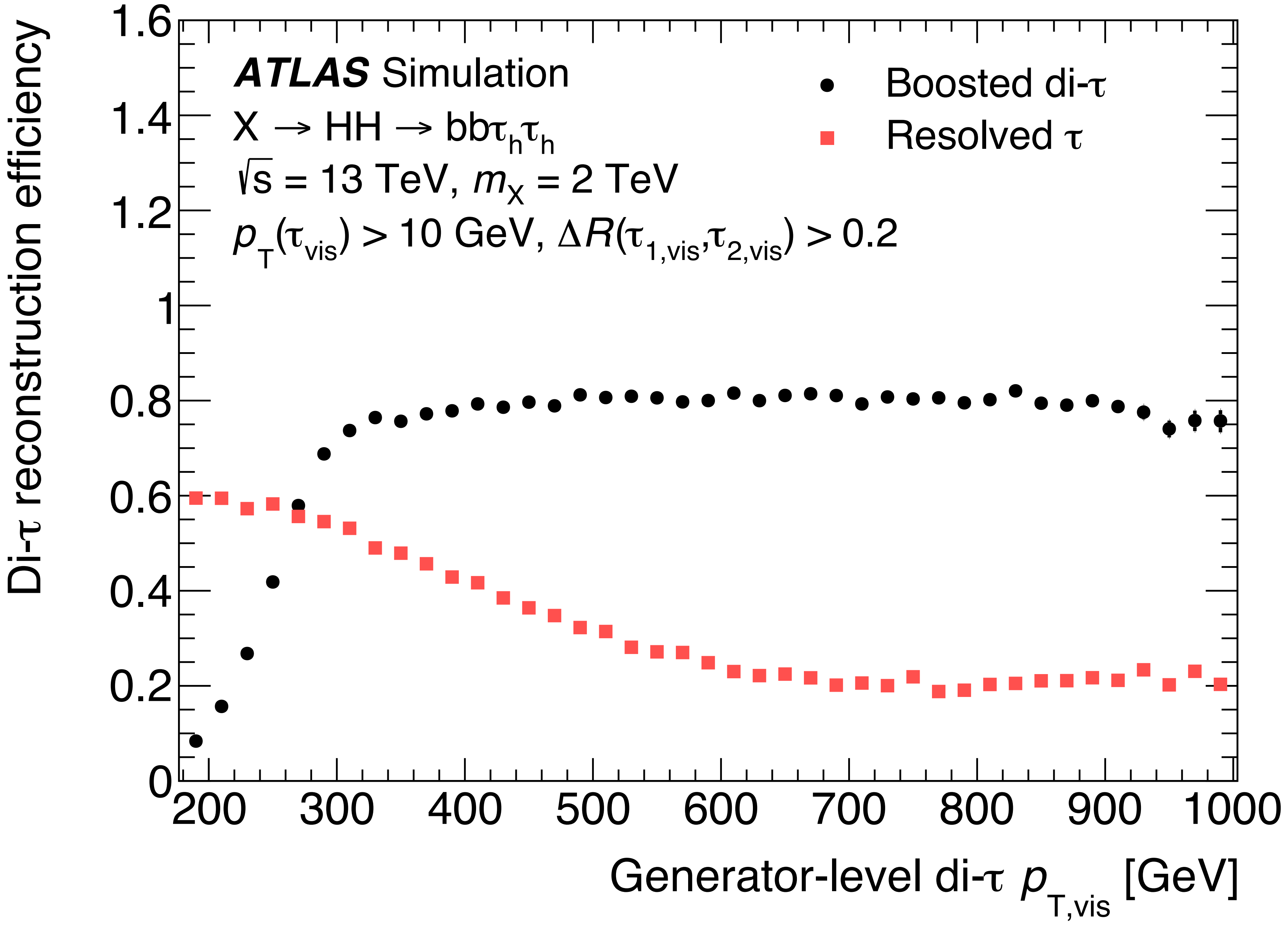
Collimated $\tau\tau$ decays mutually affect isolation criteria

Reconstruction of $\tau_h\tau_h$ as a single R =1 jet

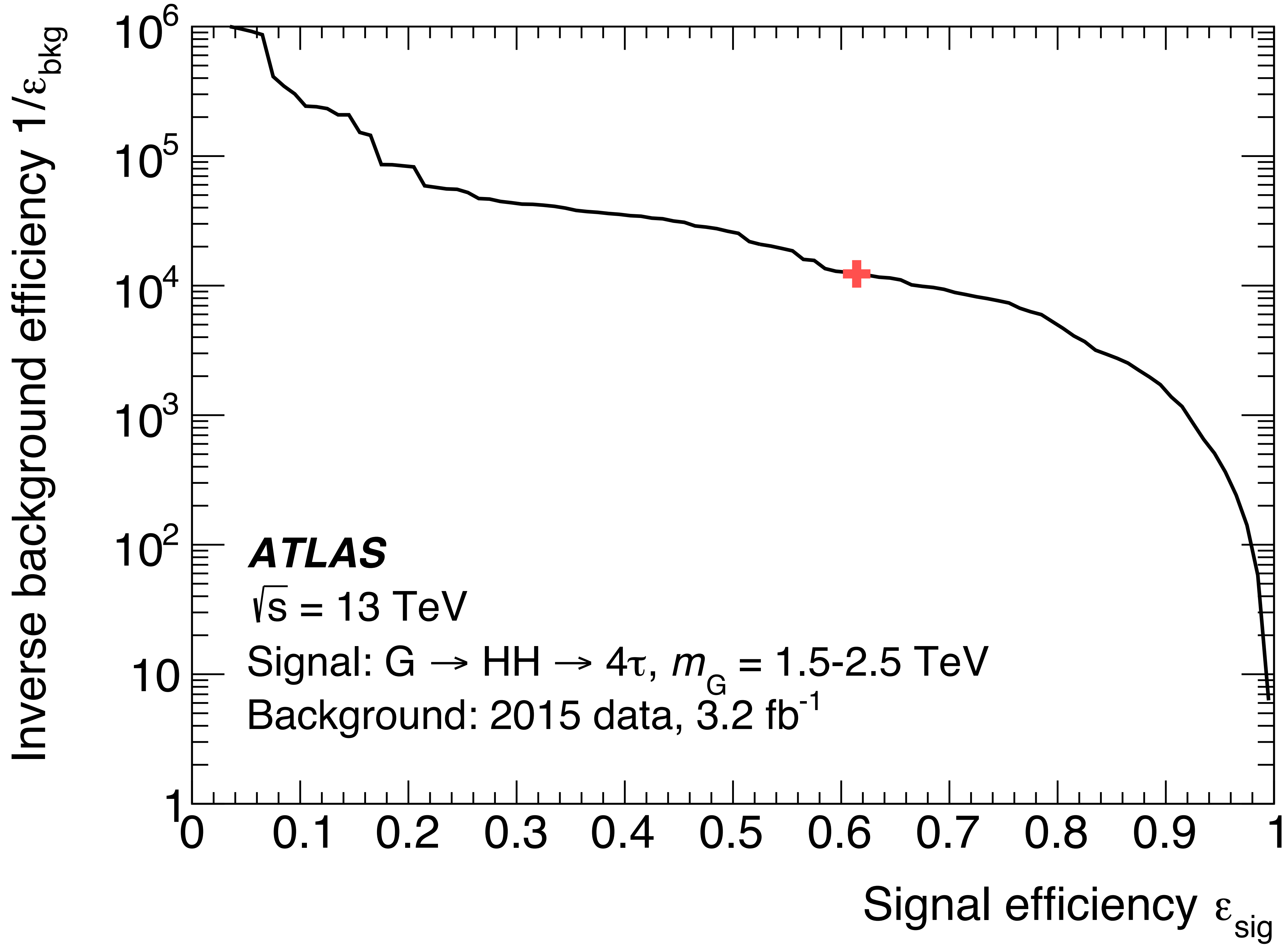


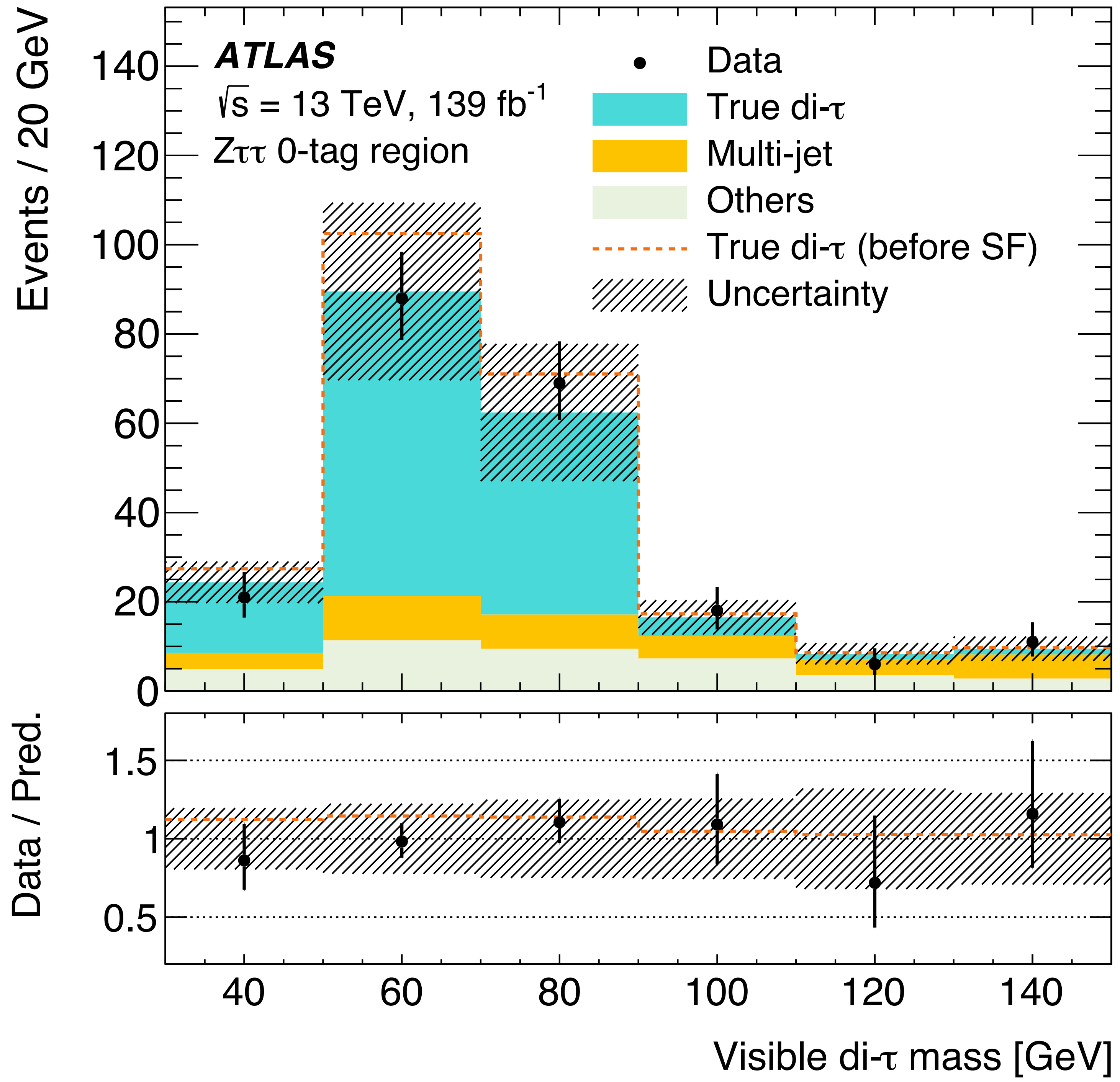
- Target $\tau_h\tau_h$ only
- Seeded by R=1 jets with $p_T > 300$ GeV
- Constituents reclustered with R = 0.2, 2 sub-jets needed
- τ_h reconstructed from tracks matched to sub-jets, residual tracks for isolation





Boosted $H \rightarrow \tau\tau$ identification





- BDT for identification
  - stable performance vs p_T and pileup
- 17 high-level features
  - multiplicity of tracks, isolation, and distribution of energy between subjets and R=1 jet
- Require 1 or 3 tracks in the subjet core to tag $\pi^\pm$ / $\pi^\pm\pi^\mp\pi^\pm$
  - 80% ε_{sig}, 5x better bkg. rejection
- Calibration with high p_T $Z\rightarrow\tau\tau$ + b jet veto

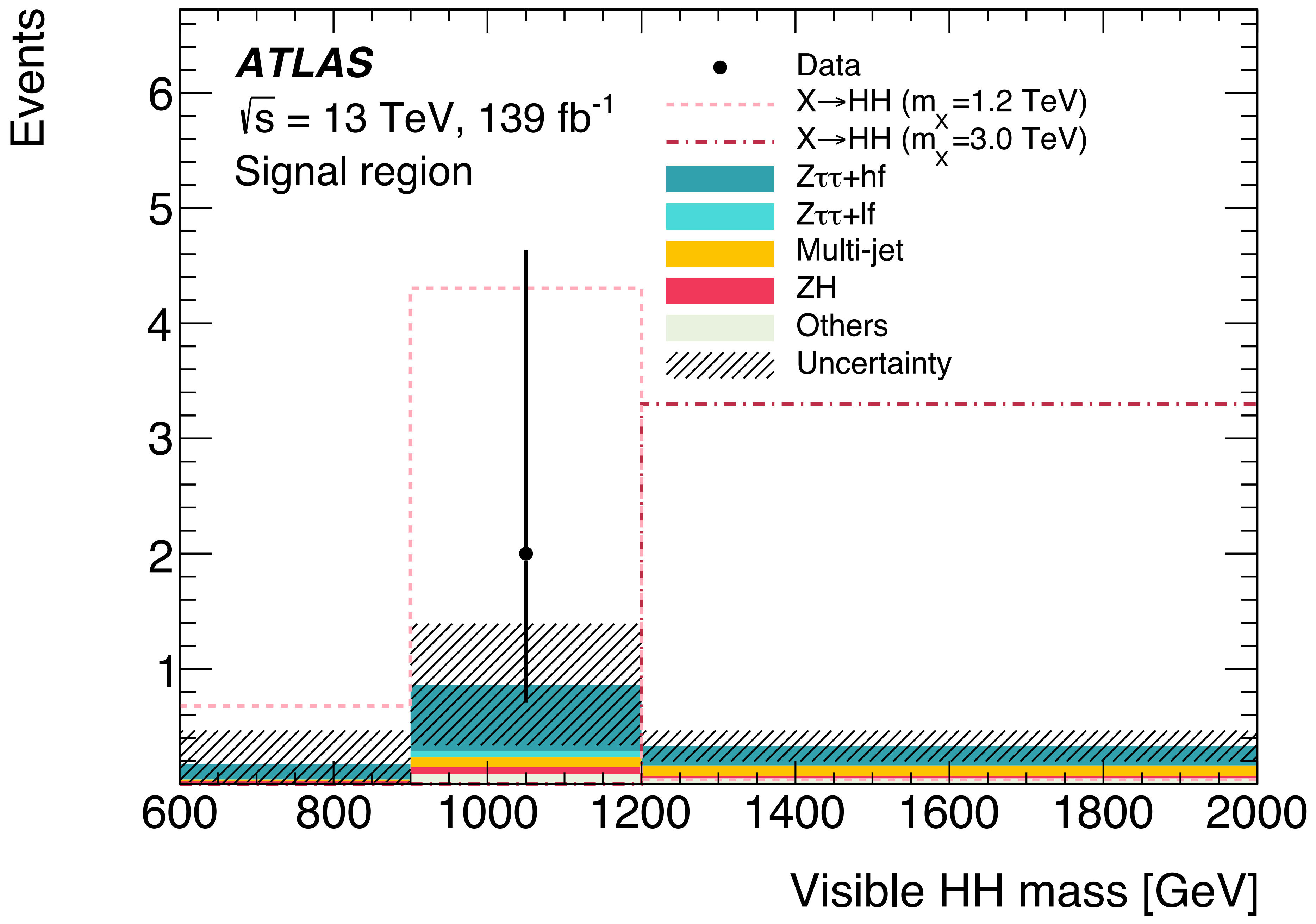
BDT-based signal identification

Tagging likely to benefit of more advanced constituent-based tagging techniques

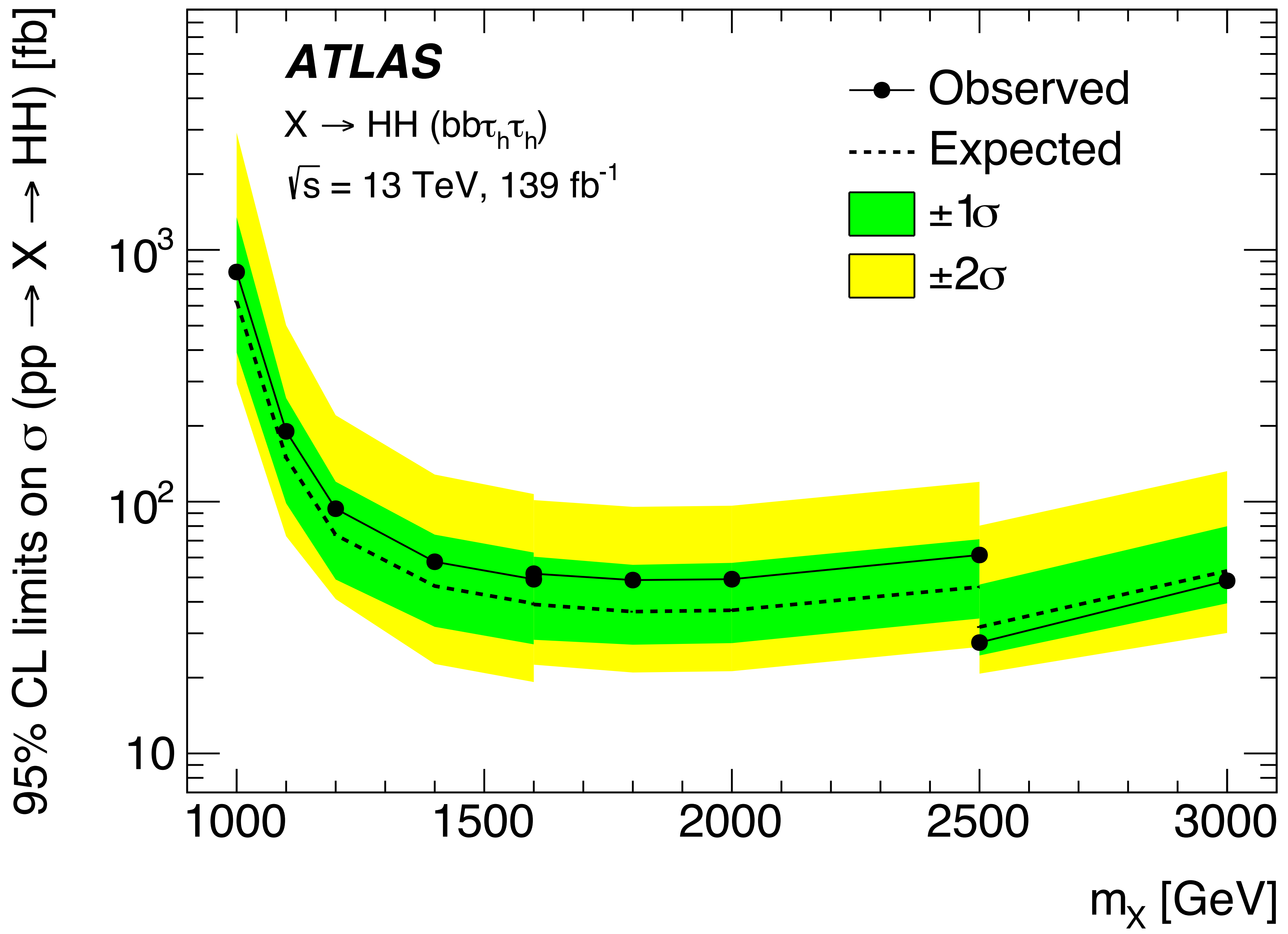
Application example : $X \to HH \to bb\tau\tau$

JHEP 11 (2020) 163

- $H(bb)H(\tau\tau)$ events reconstructed with two large radius jets
  - mass compatible with 125 GeV, 2 subjets b-tagging
- Multijet background estimated from data, $Z \to \tau\tau$ from MC



Very small number of events → simple counting experiment



Limit discontinuities related to the change in the m_{HH}^{vis} requirement depending on the m_X hypothesis

Conclusions

- Very high $p_T(H)$ final states are interesting for HH physics
  - high-sensitivity phase space corner for SM ggF, anomalous κ_{2V} coupling, high mass resonances
- Reconstruction of $p_T(H) \gtrsim 250$ GeV decays done as a single large-radius jet (R = 1)
  - re-clustering of components to identify subjets and access jet substructure
- Improvement in the performance of the H(bb) identification obtained by moving from high-level (Xbb) to constituent-based (GN2X) taggers
  - 2.5x better multijet rejection for the same signal efficiency
  - transformer architecture well suited for identification tasks
- Boosted $H(\tau\tau)$ identification thus far only on fully hadronic final states, simpler BDT algorithm based on high-level features
  - expect a similar improvement as for H(bb) by exploiting constituent information
- Highly performant boosted H identification as key to push the exploration of the high m_{HH} regime