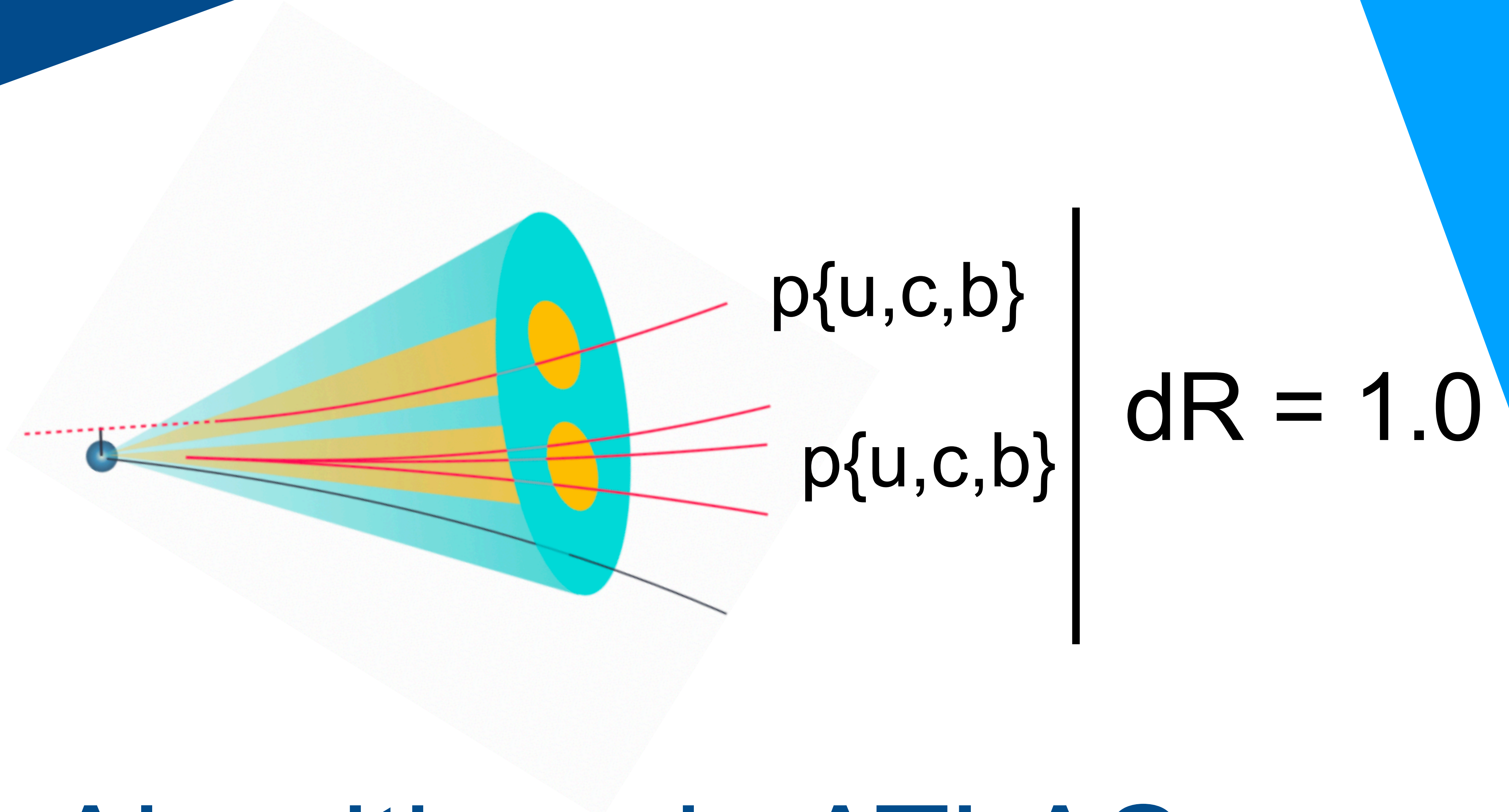


Boosted Flavour Tagging Algorithms in ATLAS

Osama Karkout

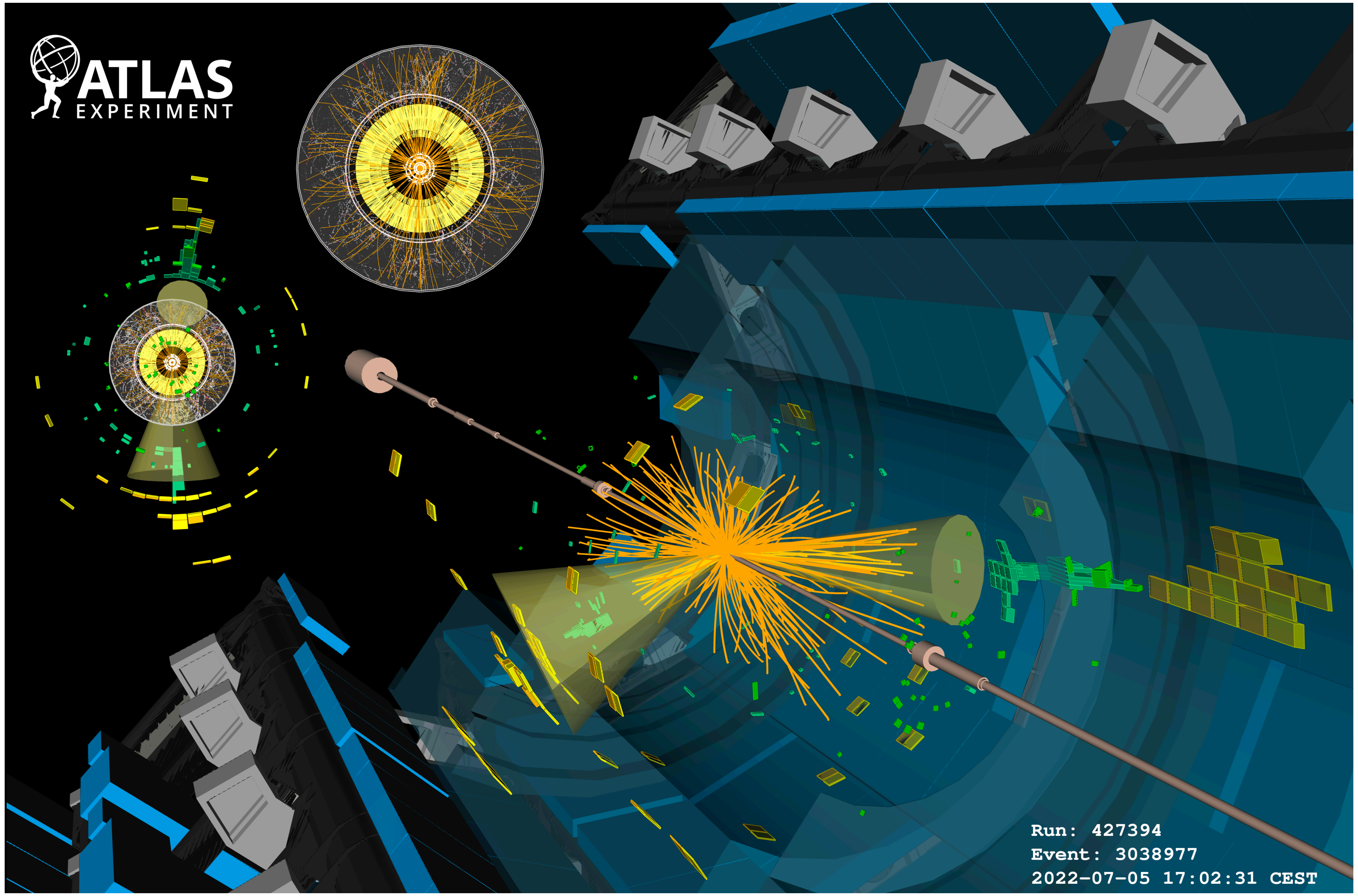
On Behalf of the ATLAS Flavour Tagging Group

Contents

1. Why boosted
2. The Xbb tagger
3. Xbb performance
4. a->bb (low mass) tagger
5. GN1 and future
6. Summary: boosted tagging is useful.

Why boosted (high transverse momentum (pt))

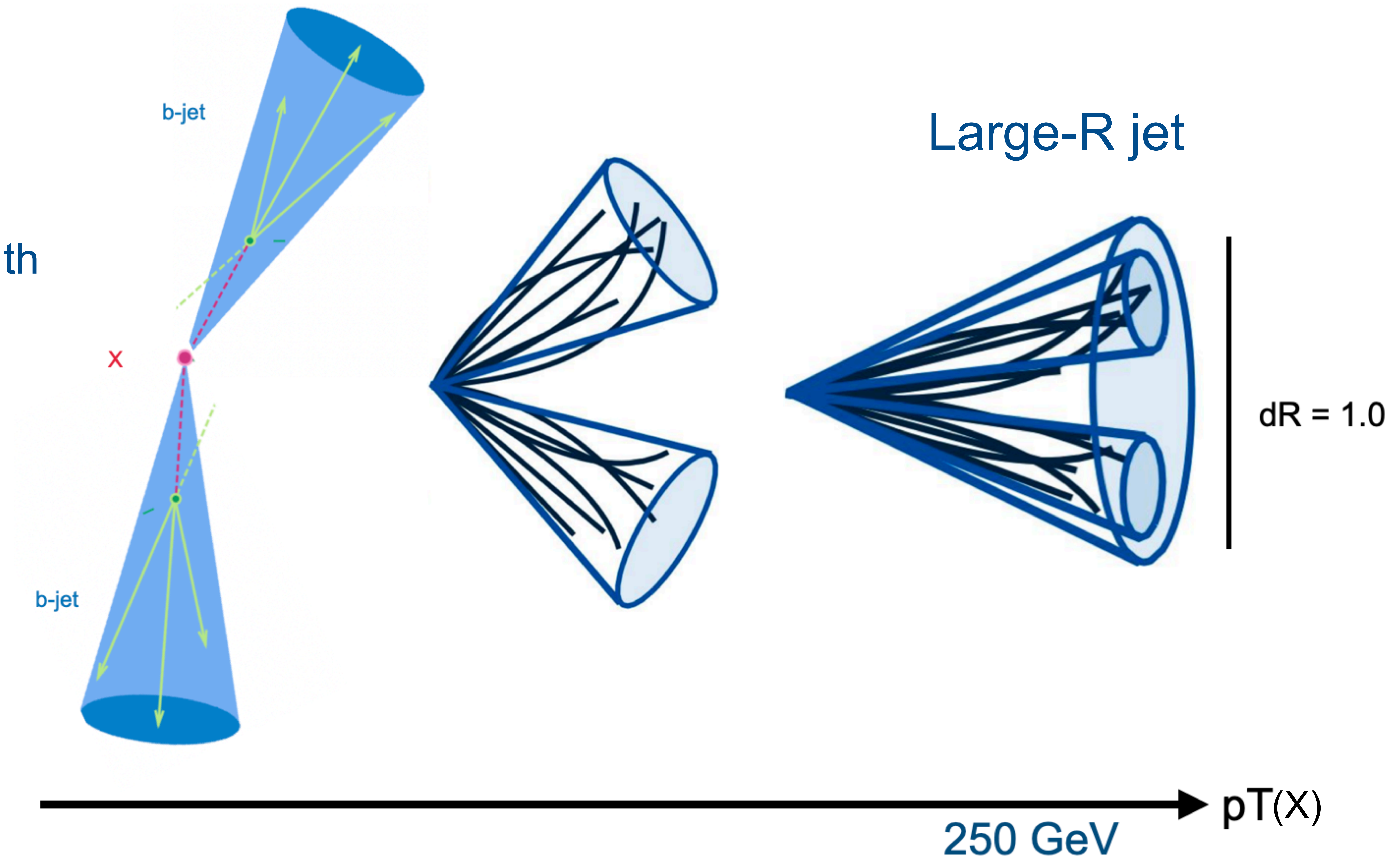
Small radius jets in ATLAS are classified according to their quark flavour: {light, c, b}



Why boosted

At high pT, small-r jets can be collimated
-> algorithms are less sensitive

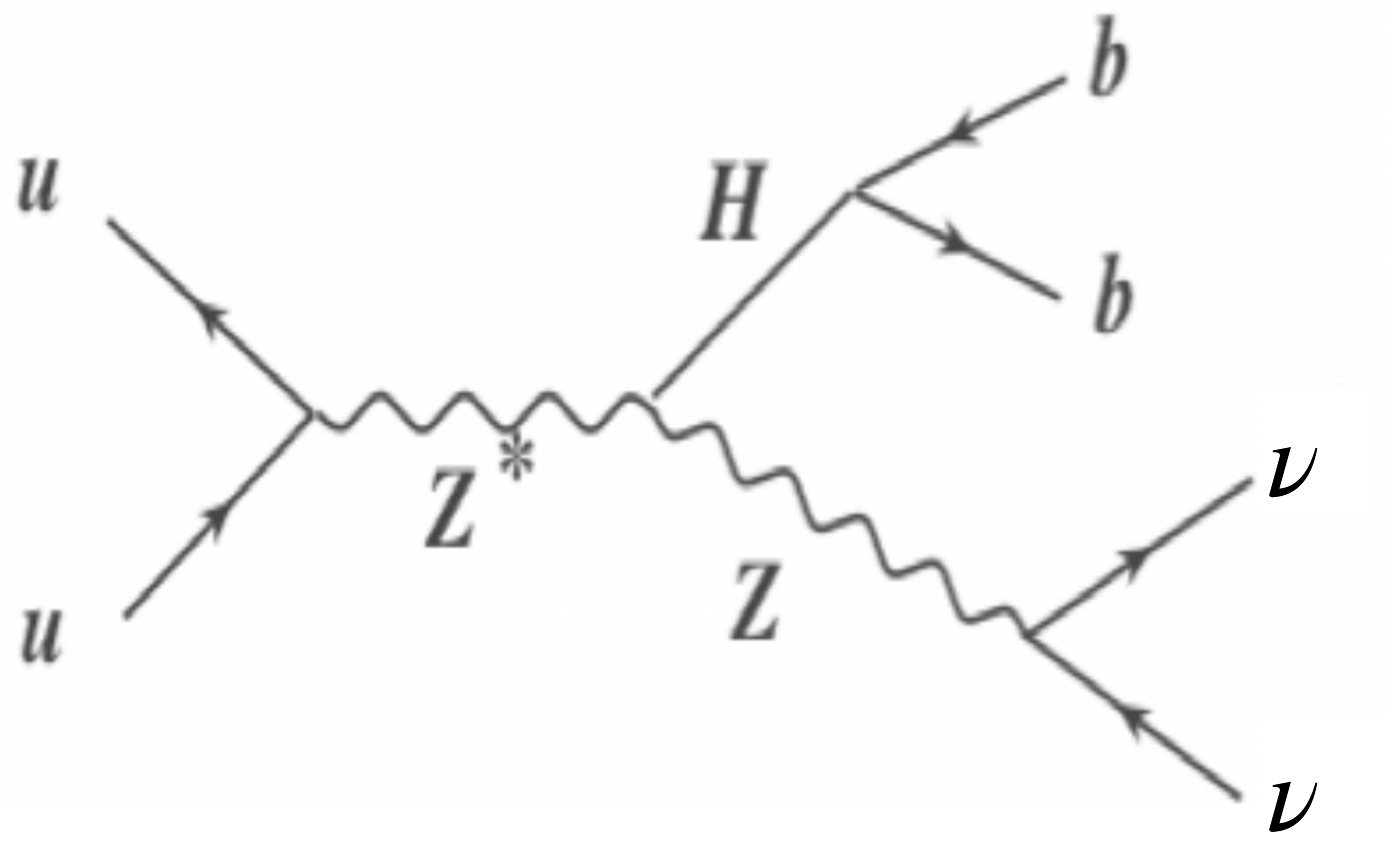
We can construct large-R jet with decay products of resonance particle



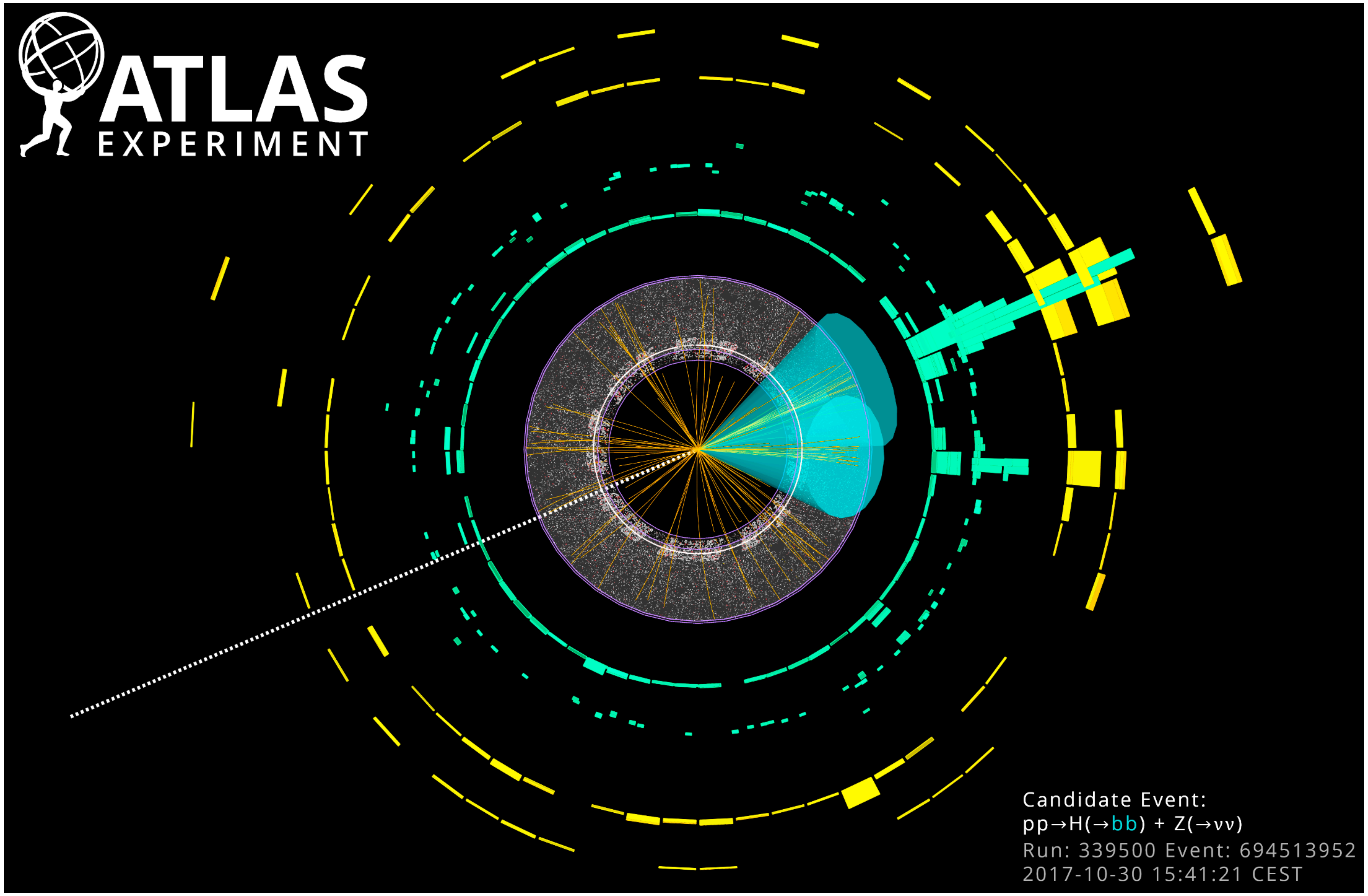
Why boosted

Boosted particles decaying hadronically
Form Large-R jets

Happens for pT(H) >~ 200 GeV
Useful for HHH?



H->bb: 58% of the time



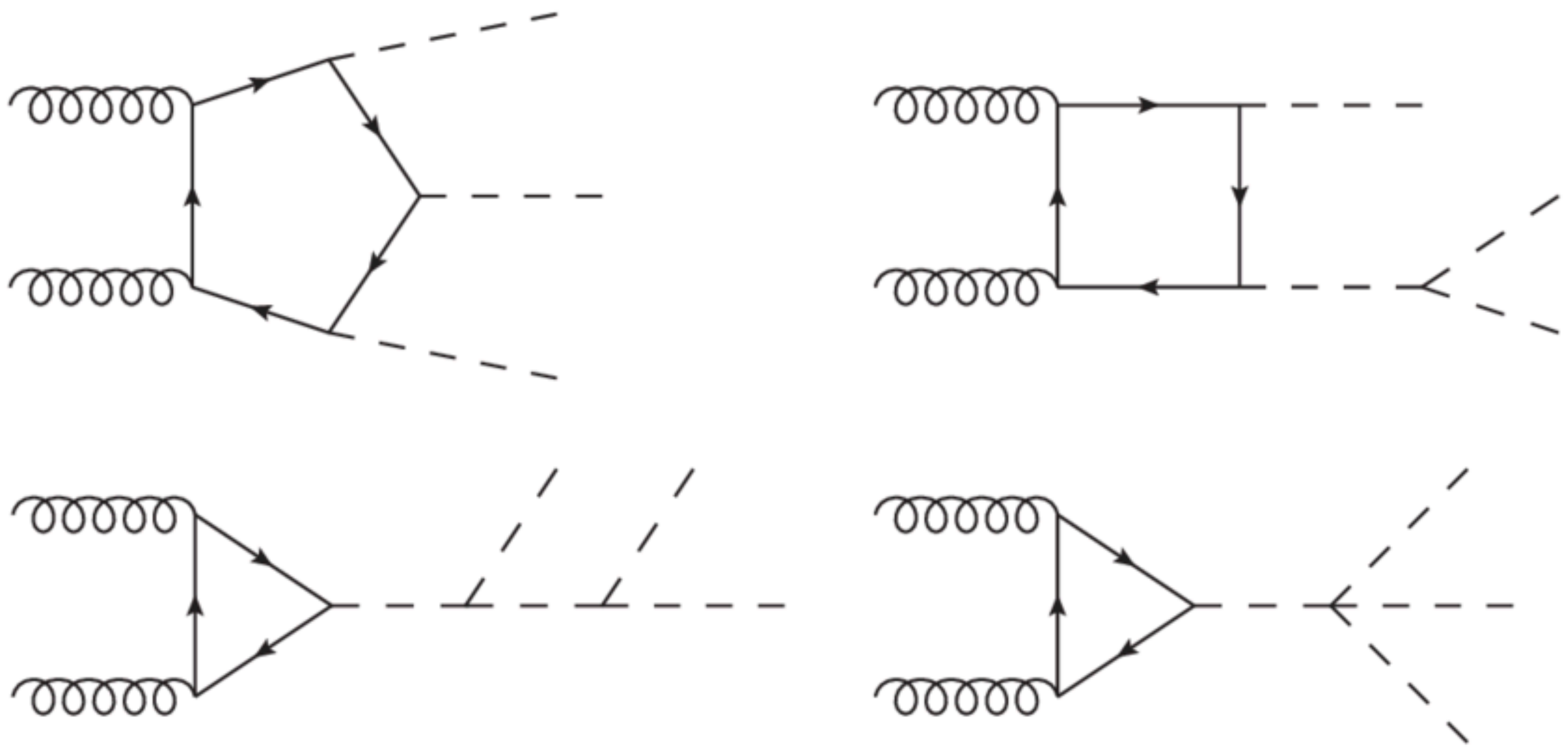
Why boosted (for SM HHH)

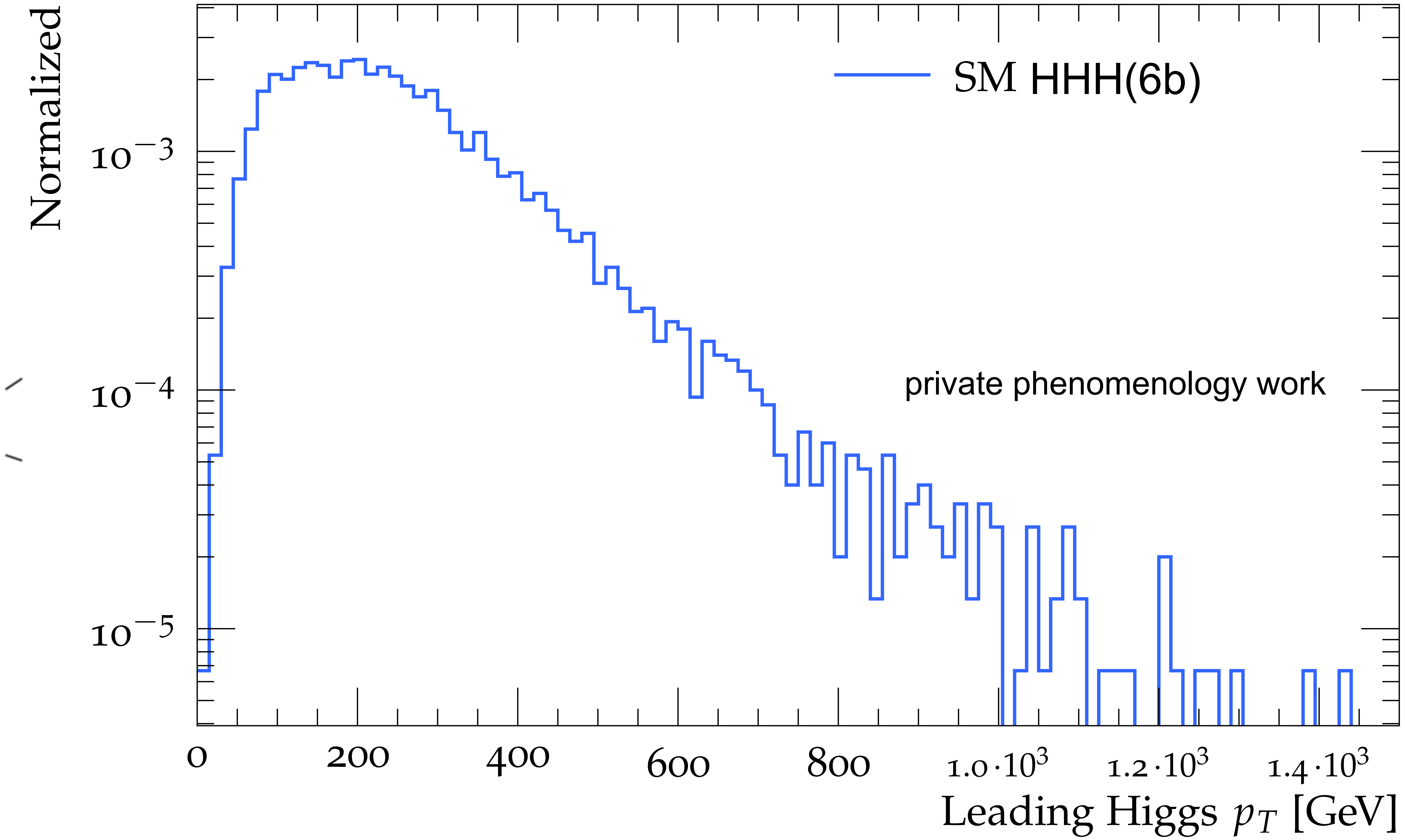
Boosted tagging for HHH in the SM could be useful!

Plot below: SM pp -> HHH production at 13.6 TeV, 6b channel
Central b-jet must contain a truth-level B-hadron and have
$p_T > 20\ GeV,\ \ |\eta| < 2.5$
preselection: events with at least 4 central b-jets.
with paring algorithm to reconstruct H

Left plot:

Rivet generated distribution of leading Higgs pT distribution.
Large part of the distribution is boosted!
Plot by Pim Bijl and Carlo Pandini

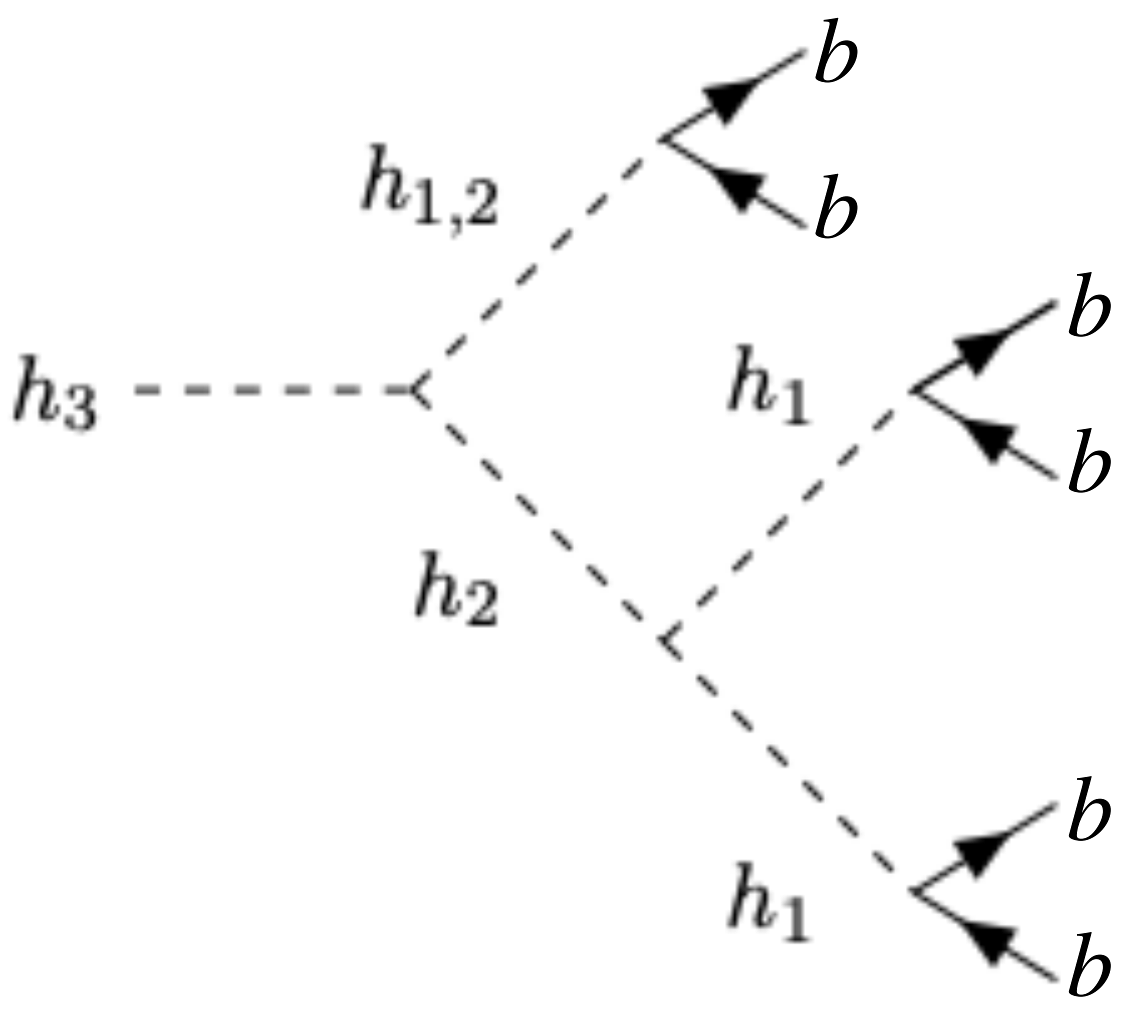




Why boosted (for BSM HHH)

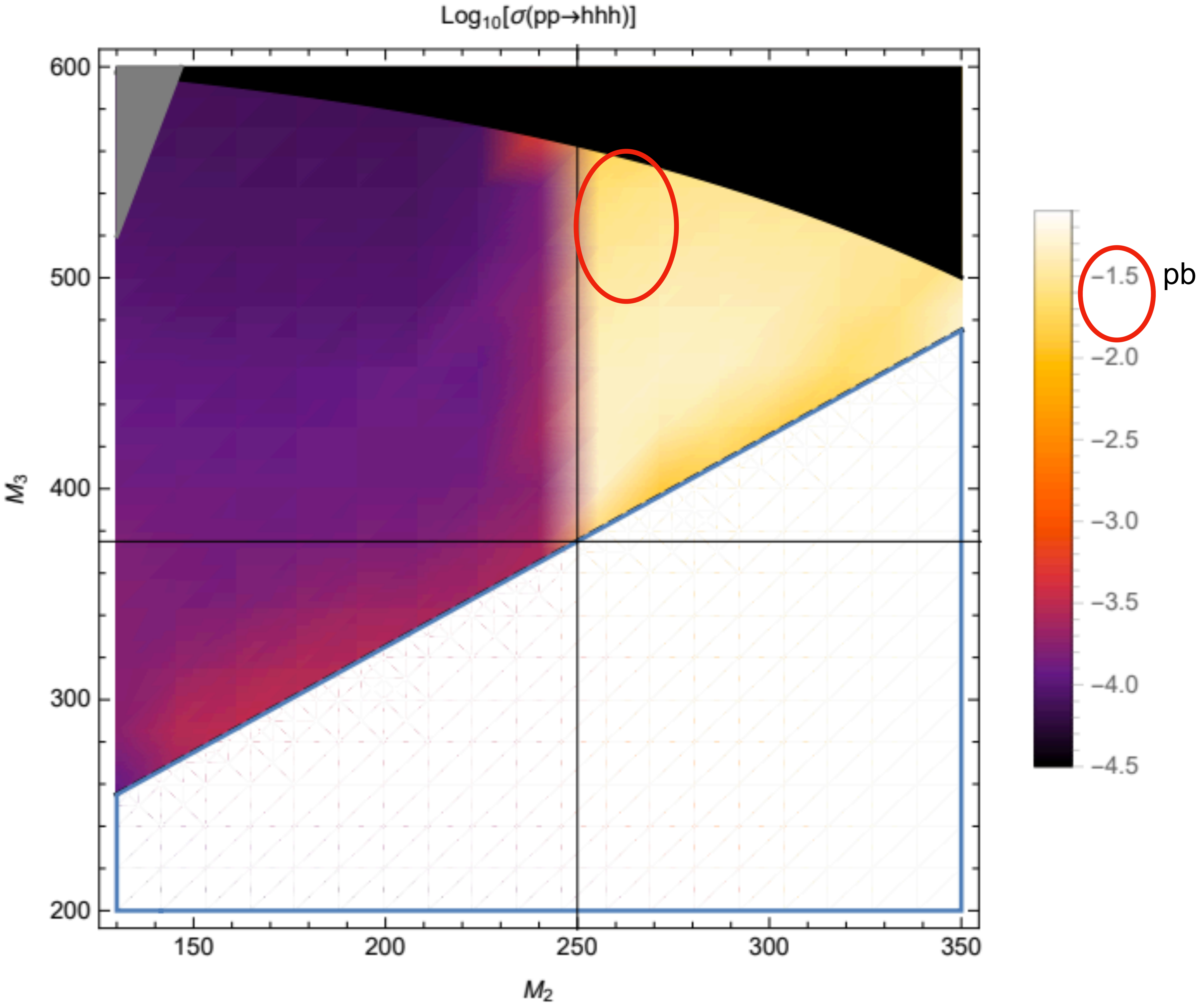
Example: TRSM: SM + two singlets
h1 can be our Higgs of 125 GeV

Boosted if M3 > M2 + 125 GeV
but xsec ~ 50 fb (similar to HH production in SM)



$$V = \mu_\Phi^2 \Phi^\dagger \Phi + \lambda_\Phi (\Phi^\dagger \Phi)^2 + \mu_S^2 S^2 + \lambda_S S^4 + \mu_X^2 X^2 + \lambda_X X^4 + \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2 + \lambda_{SX} S^2 X^2 \,.$$

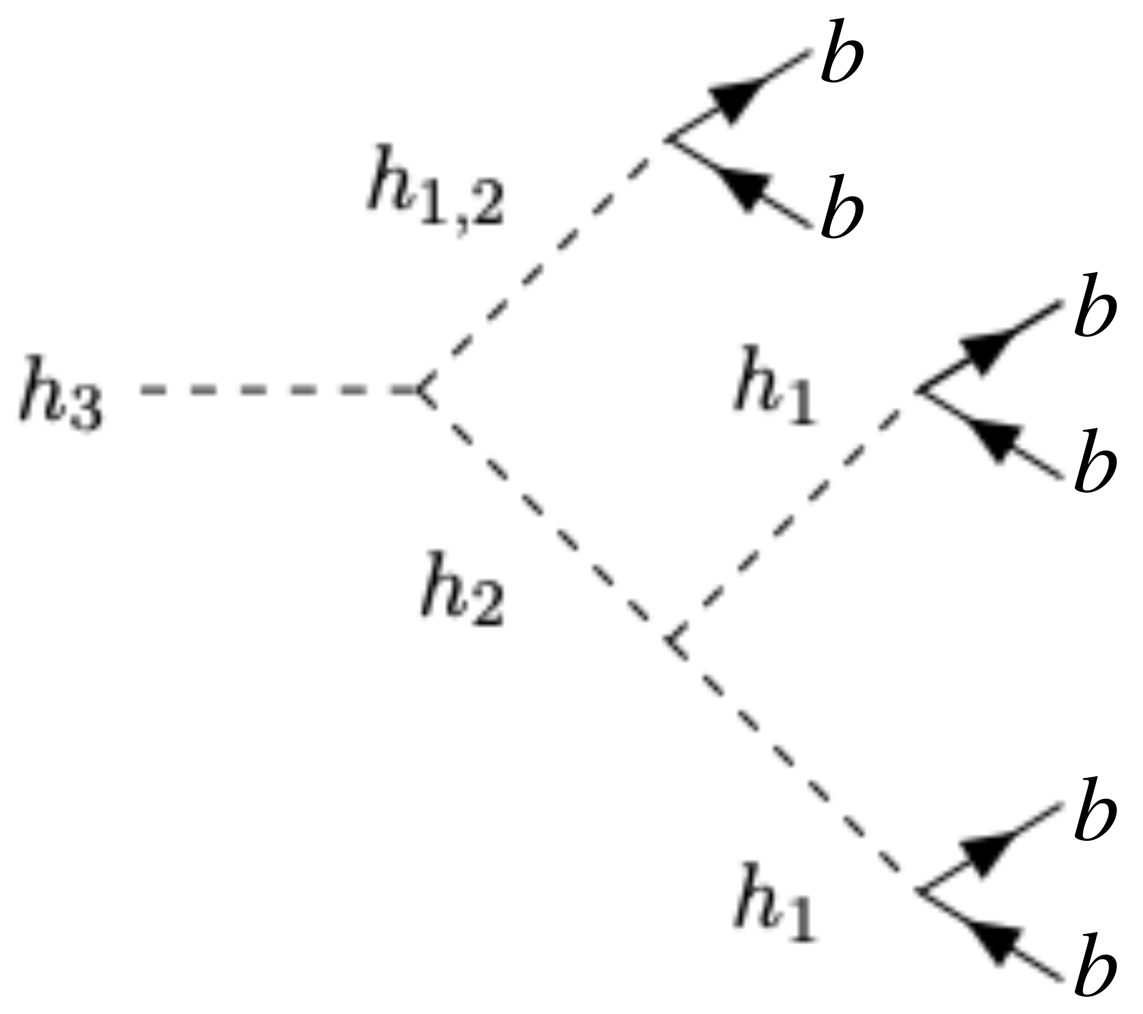
TRSM BP3 @14TeV:



Why boosted (for BSM HHH)

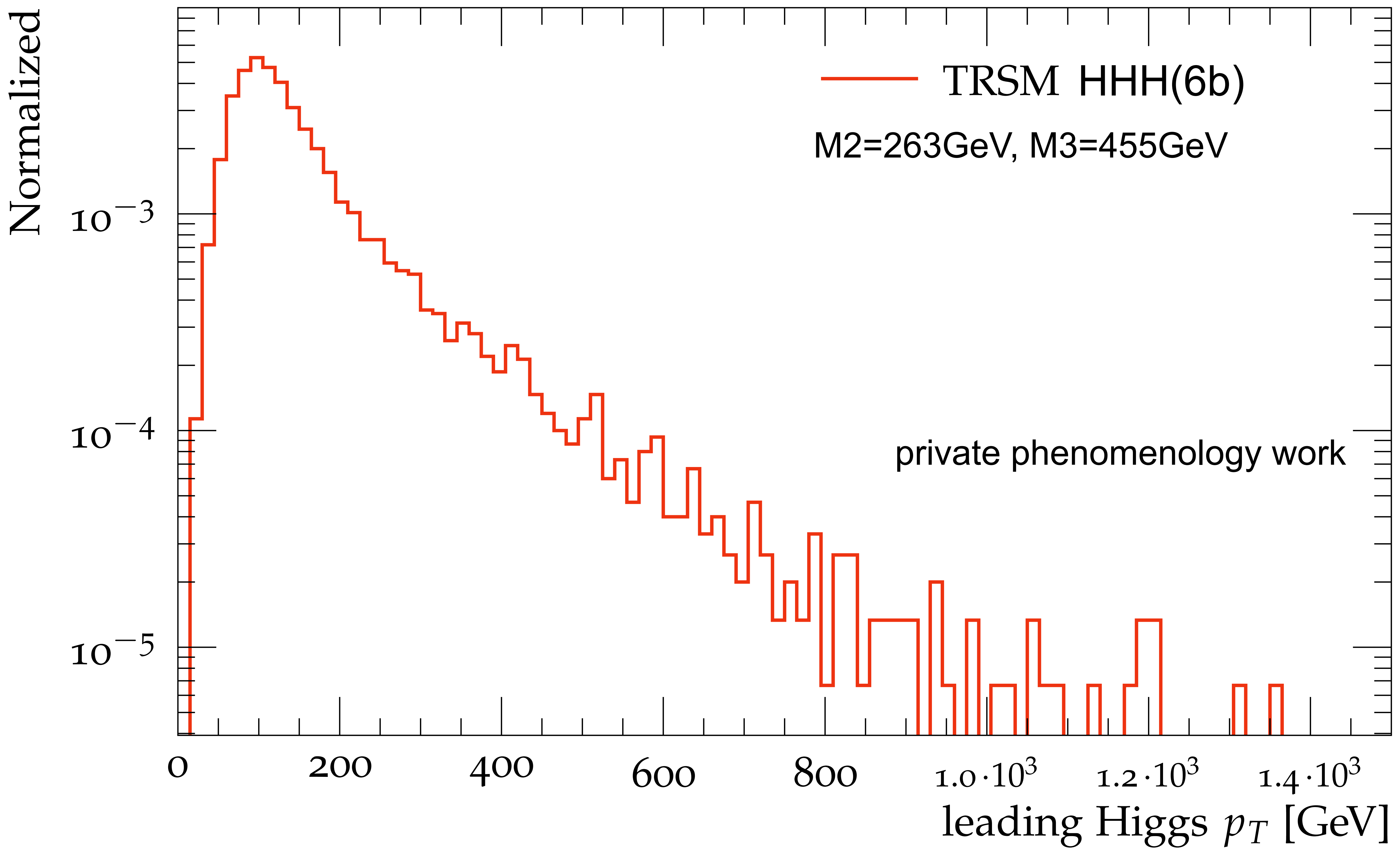
Example: TRSM: SM + two singlets
h1 can be our Higgs of 125 GeV

Left plot: Rivet generated distribution of leading Higgs pT distribution.
Large part of the distribution is boosted!
Plot by Pim Bijl and Carlo Pandini
xsec ~ 50 fb (similar to HH production)



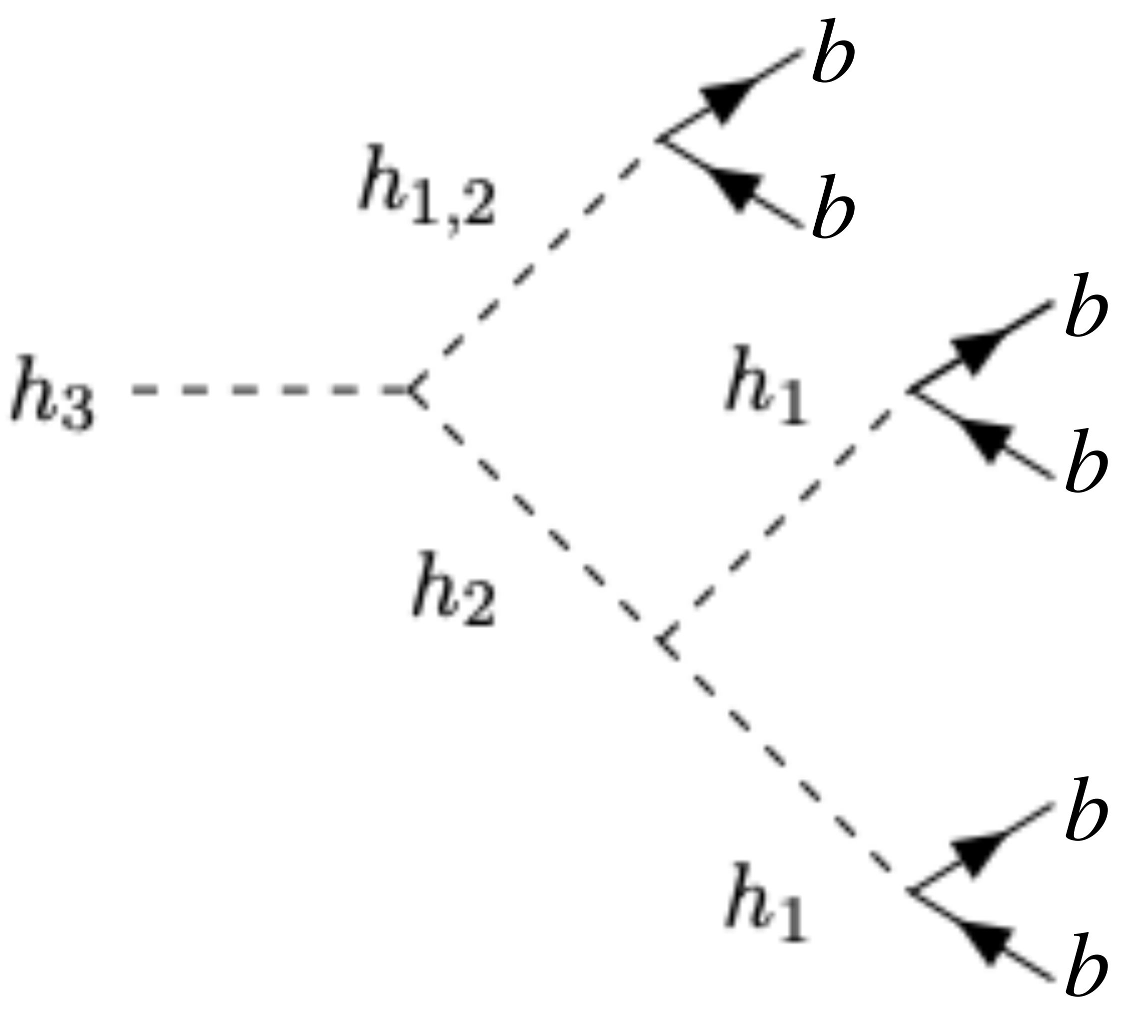
$$V = \mu_\Phi^2 \Phi^\dagger \Phi + \lambda_\Phi (\Phi^\dagger \Phi)^2 + \mu_S^2 S^2 + \lambda_S S^4 + \mu_X^2 X^2 + \lambda_X X^4 + \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2 + \lambda_{SX} S^2 X^2 .$$

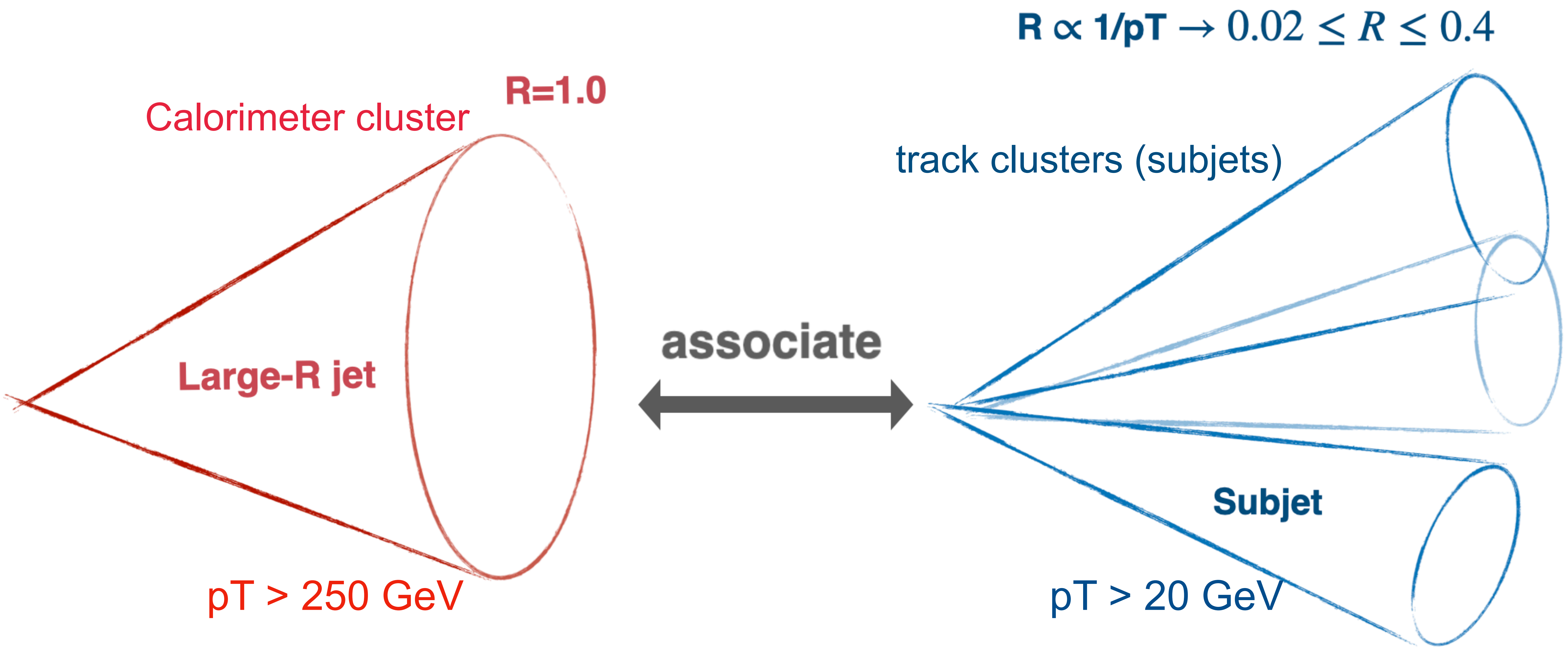
Plot below: TRSM BP3 (M2=263 GeV, M3=455 GeV)
pp -> HHH production at 13.6 TeV, 6b channel
Central b-jet must contain a truth-level B-hadron and have
$p_T > 20\ GeV, \ |\eta| < 2.5$
preselection: events with at least 4 central b-jets.
with paring algorithm to reconstruct H



Xbb Tagger (in our case, X = Higgs)

We construct Large-R jet
with 3 or fewer subjets



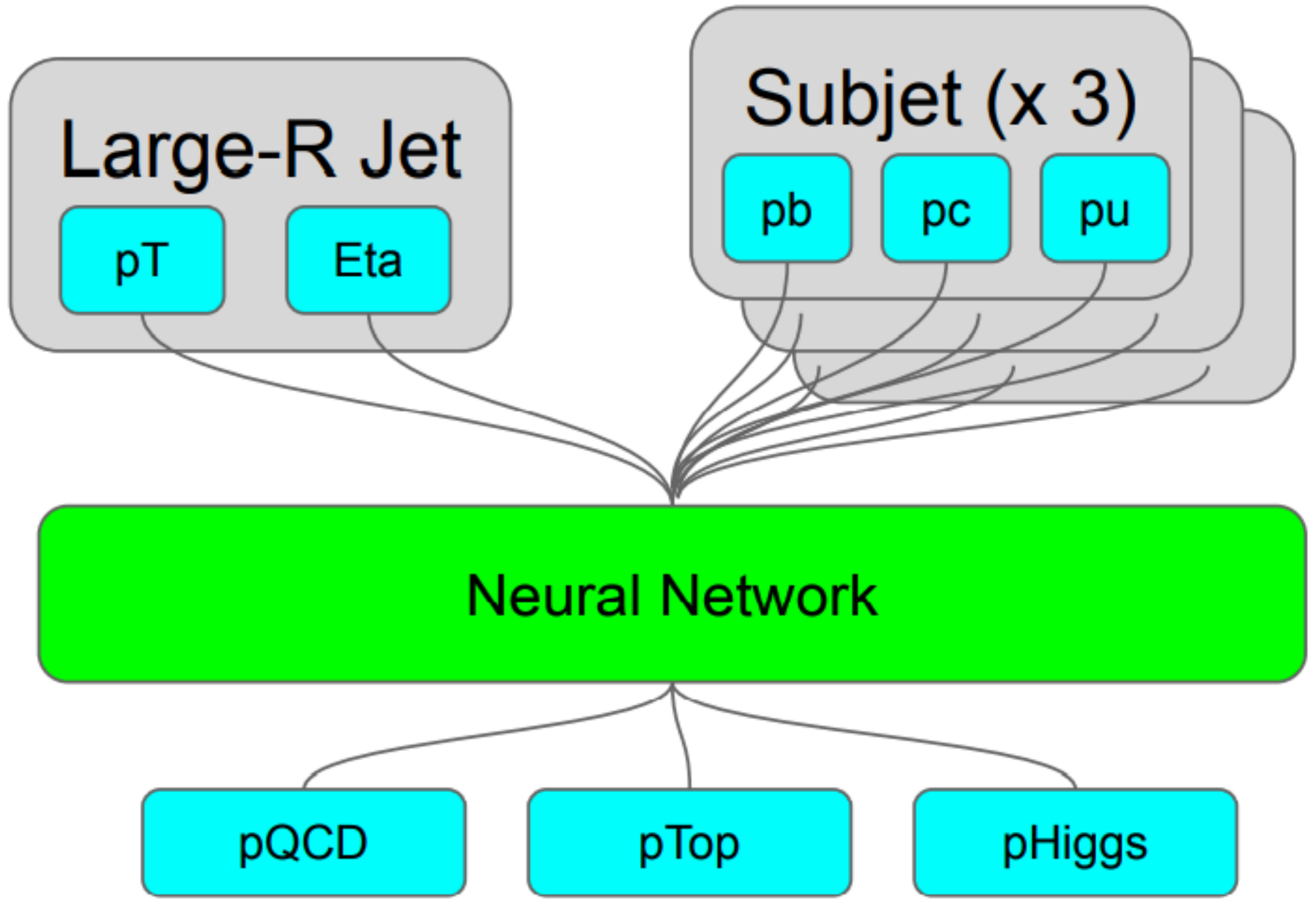


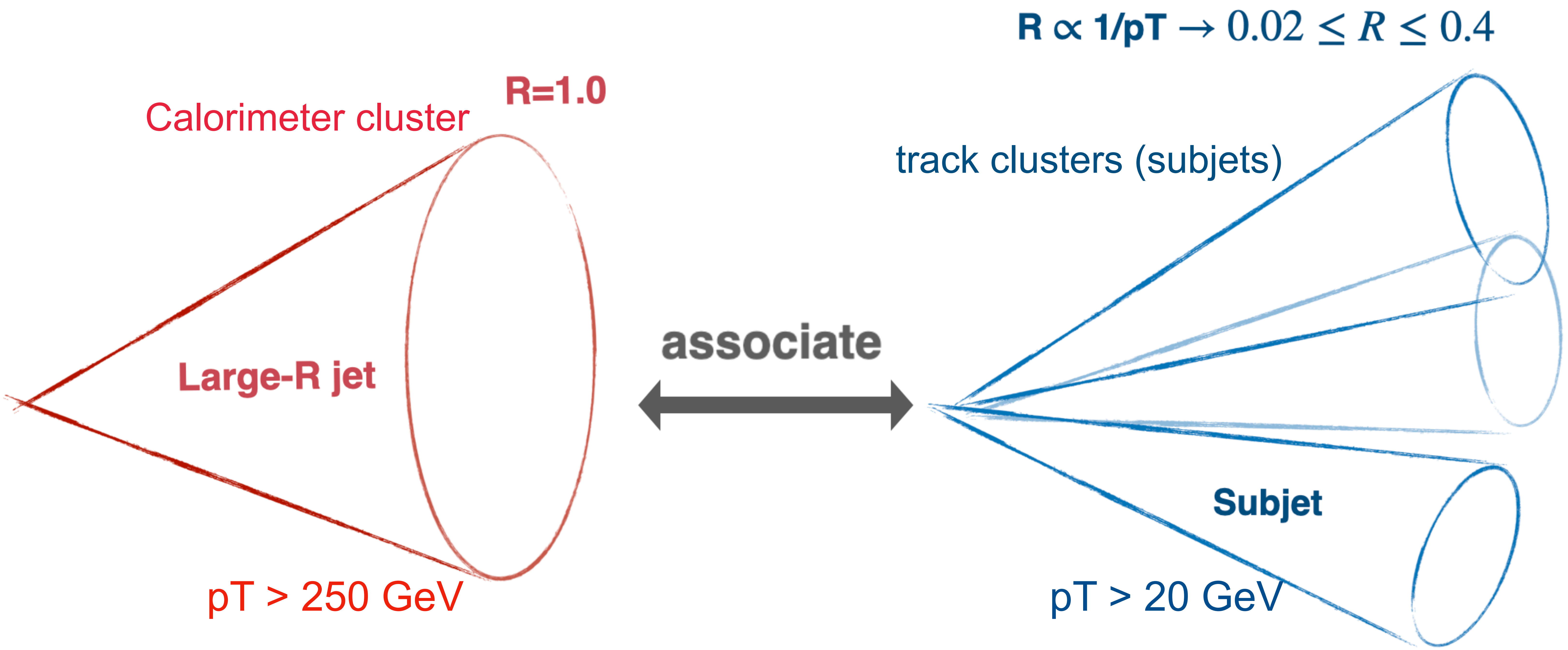
https://cds.cern.ch/record/2724739/files/ATL-PHYS-PUB-2020-019.pdf

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2017-04/

Xbb Tagger: Feed-forward Deep Neural Network

Run-2 tagger

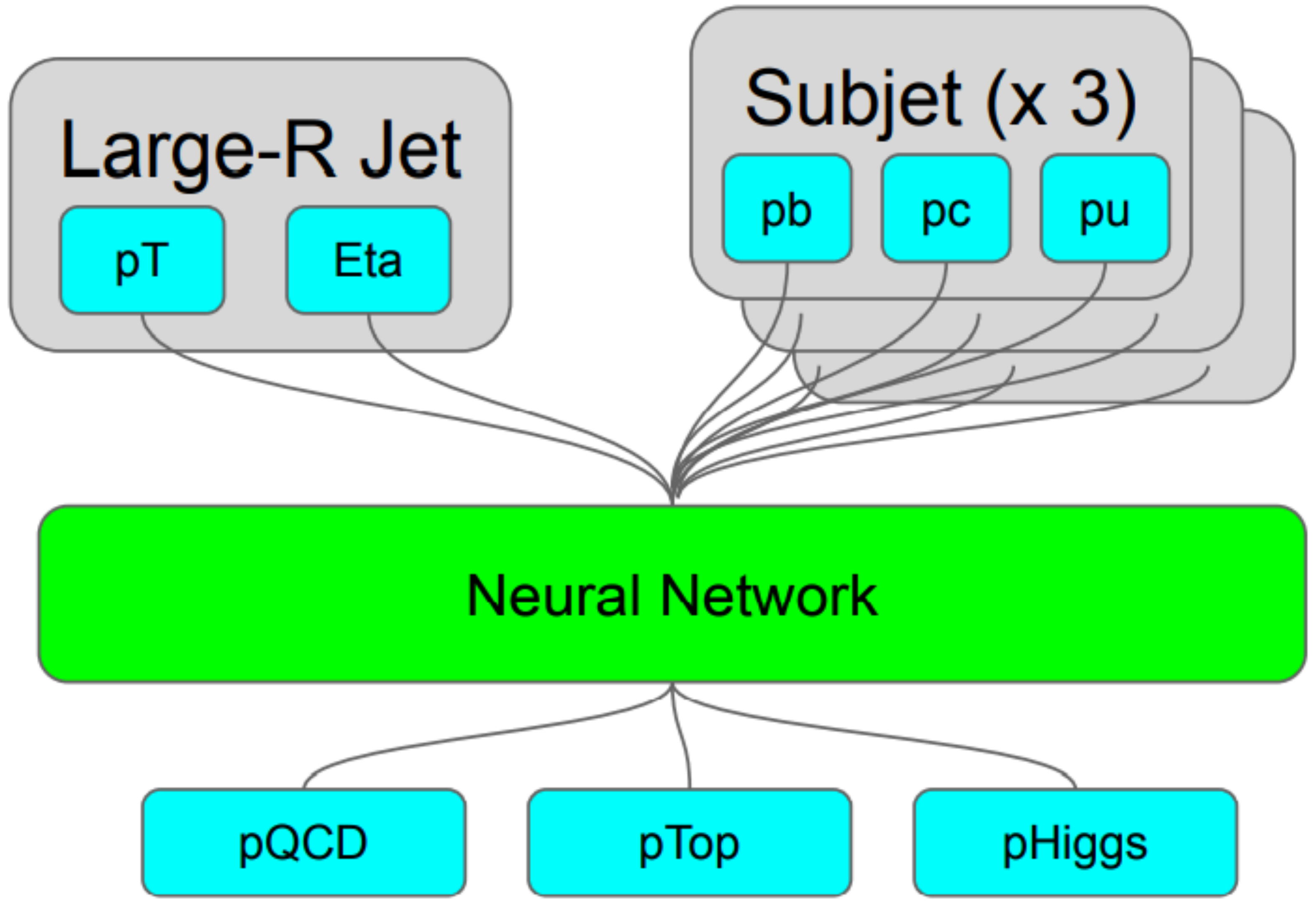




Use up to 3 subjets tagged with DL1r (deep learning algo)
-> probabilities of being {b, c, light (u)} for each subjet

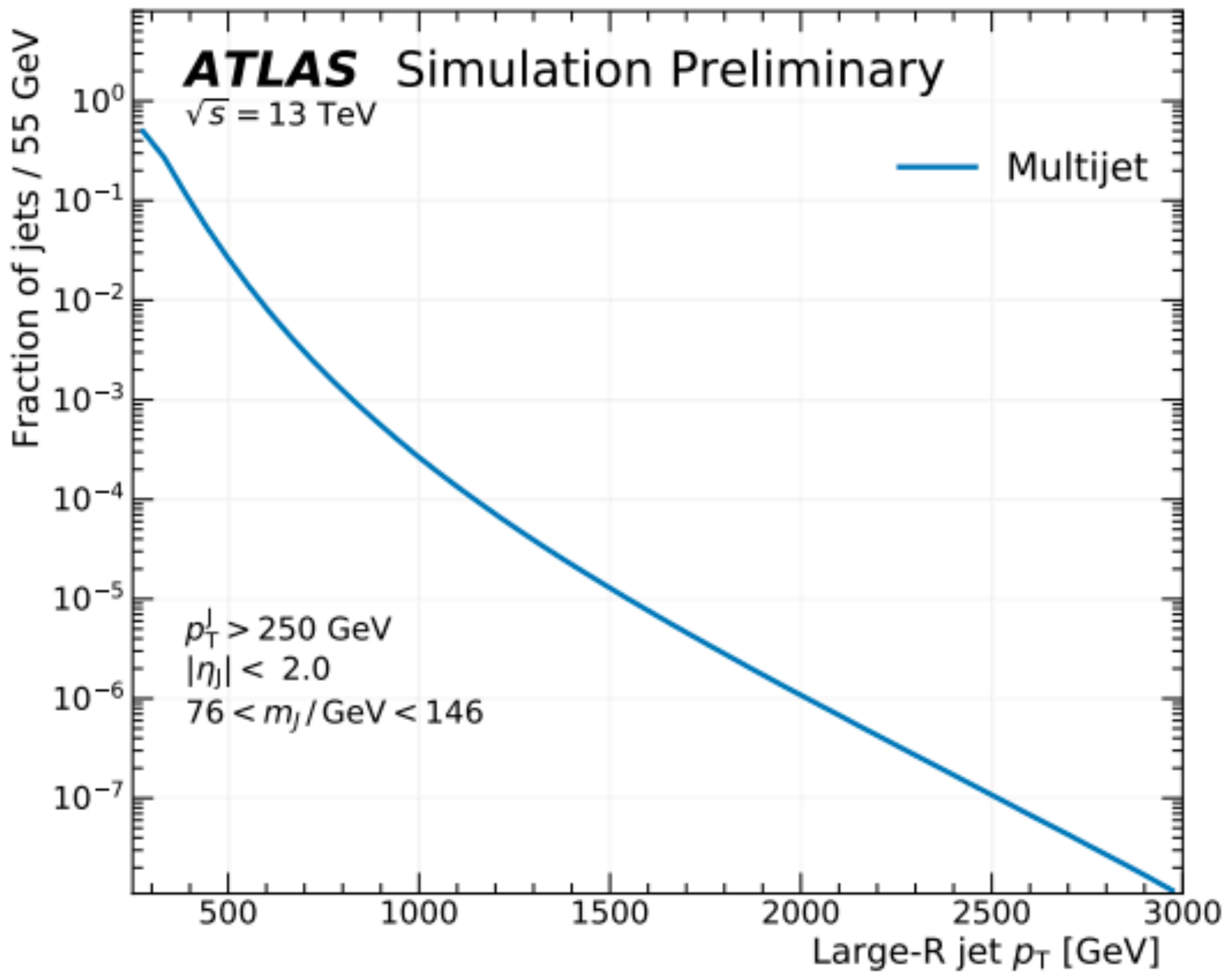
Xbb Tagger: training

Run-2 tagger



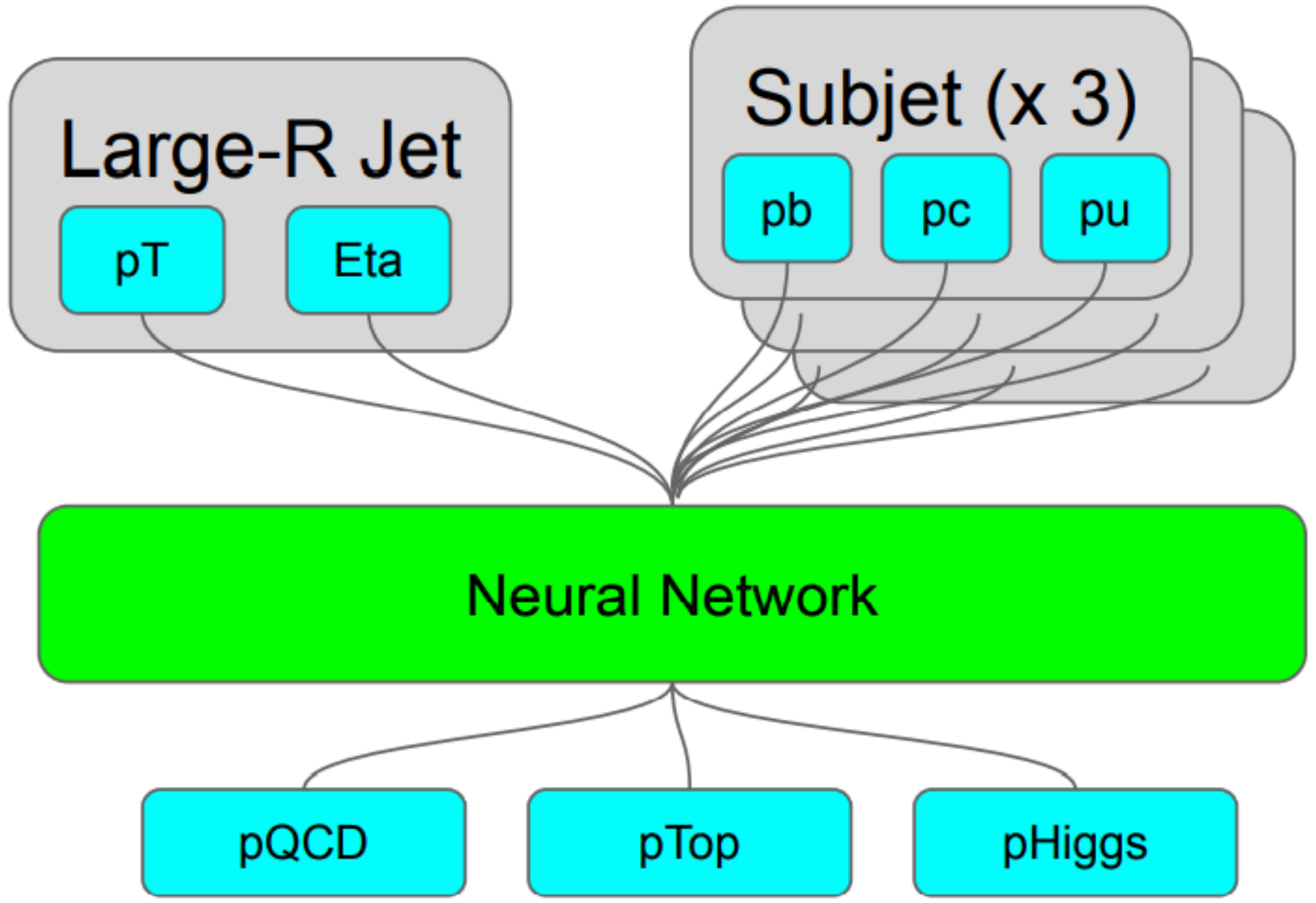
learns on MC samples:
Signal:
Jets from H->bb decays, H comes from Randall-Sundrum gravitons.
RG graviton mass: 300 GeV to 3 TeV.

Background:
Jets form:
multijet QCD samples: pt range: 250 GeV to 3000 GeV
Z'->tt: Z' mass: 750 GeV to 5 TeV

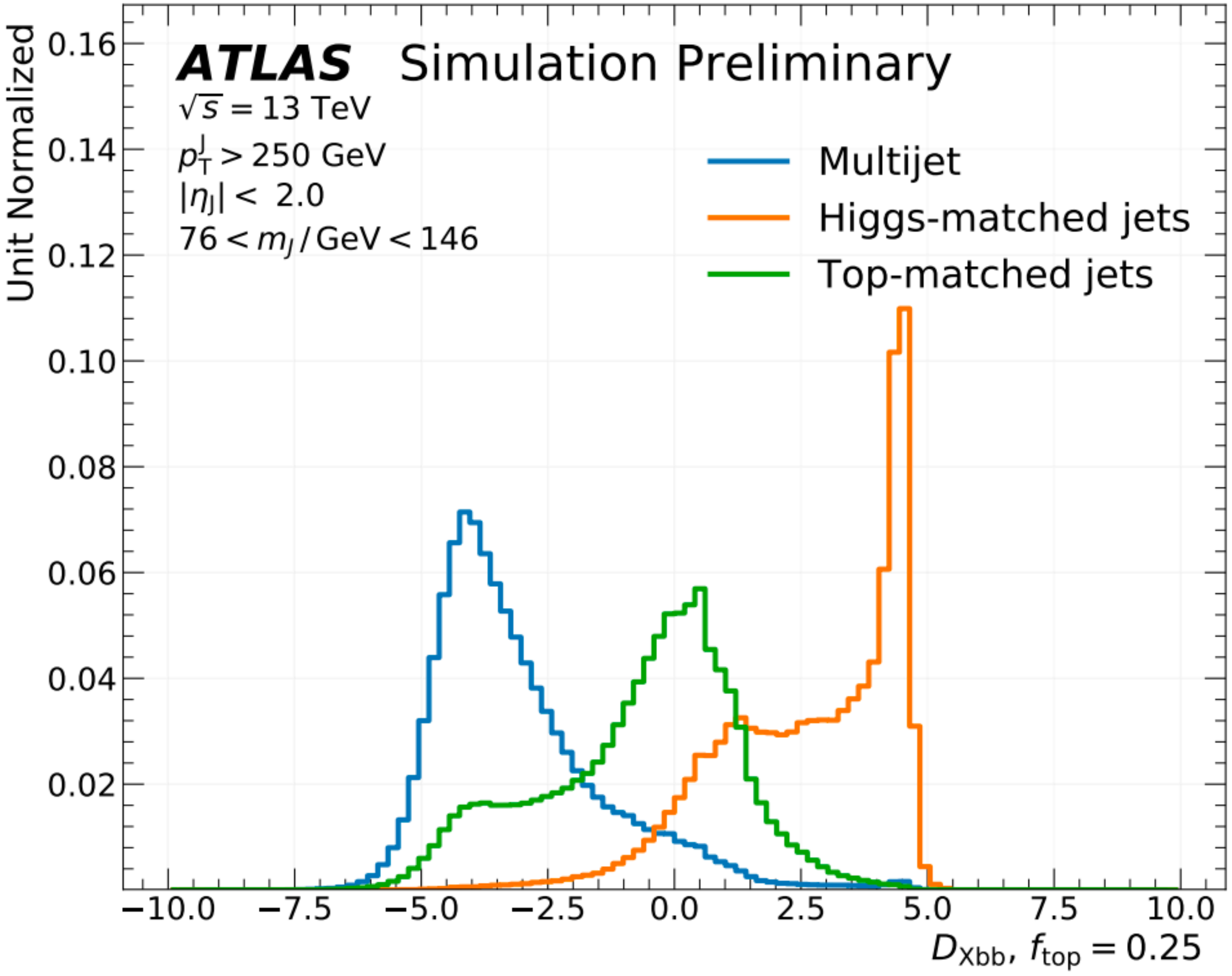


Xbb Tagger: performance

Run-2 tagger



$$D_{Xbb} = \ln \frac{p_H}{f_{\text{top}} \cdot p_t + (1 - f_{\text{top}}) \cdot p_{\text{QCD}}}$$



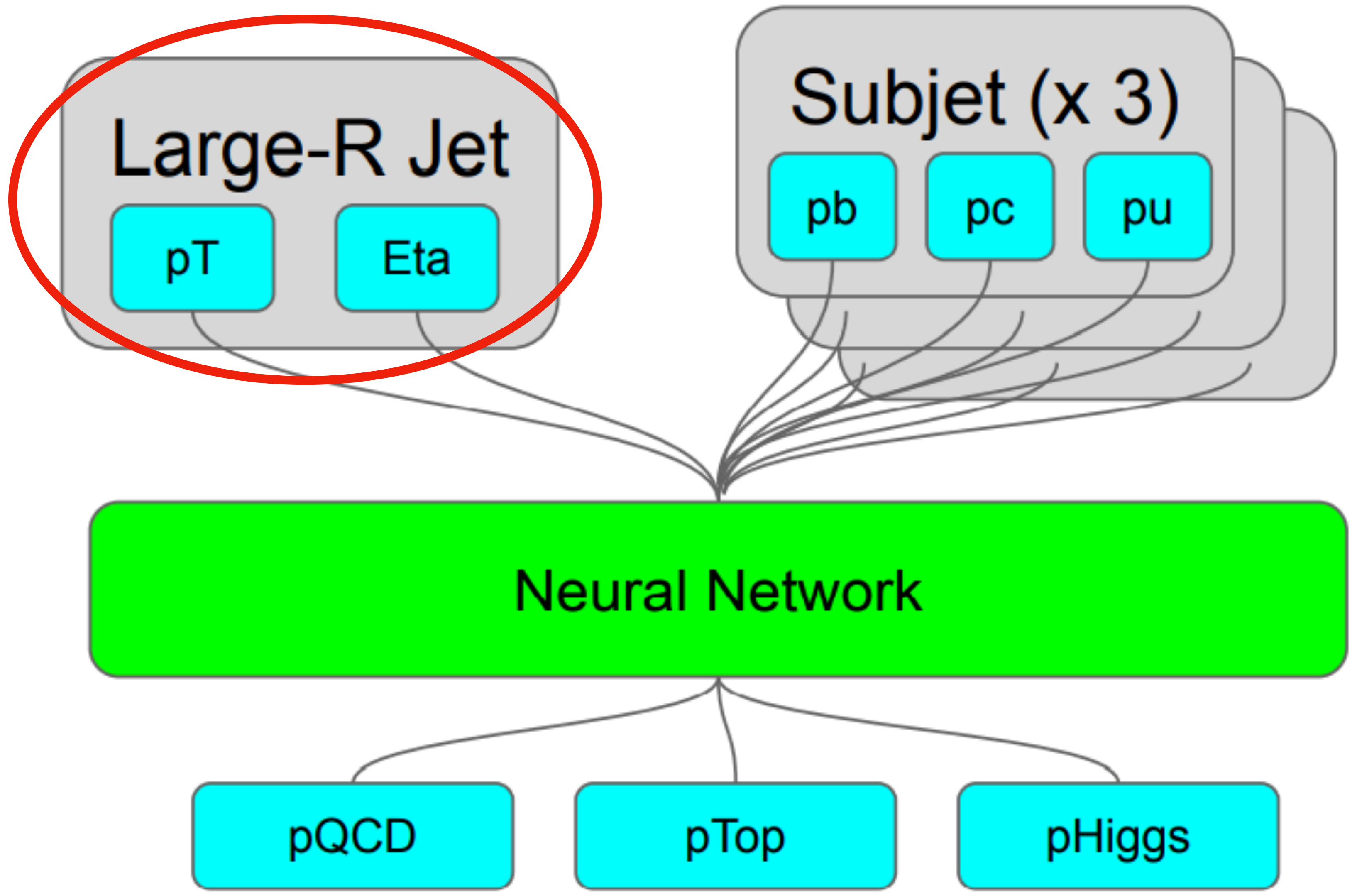
Xbb Tagger: performance

Run-2 tagger

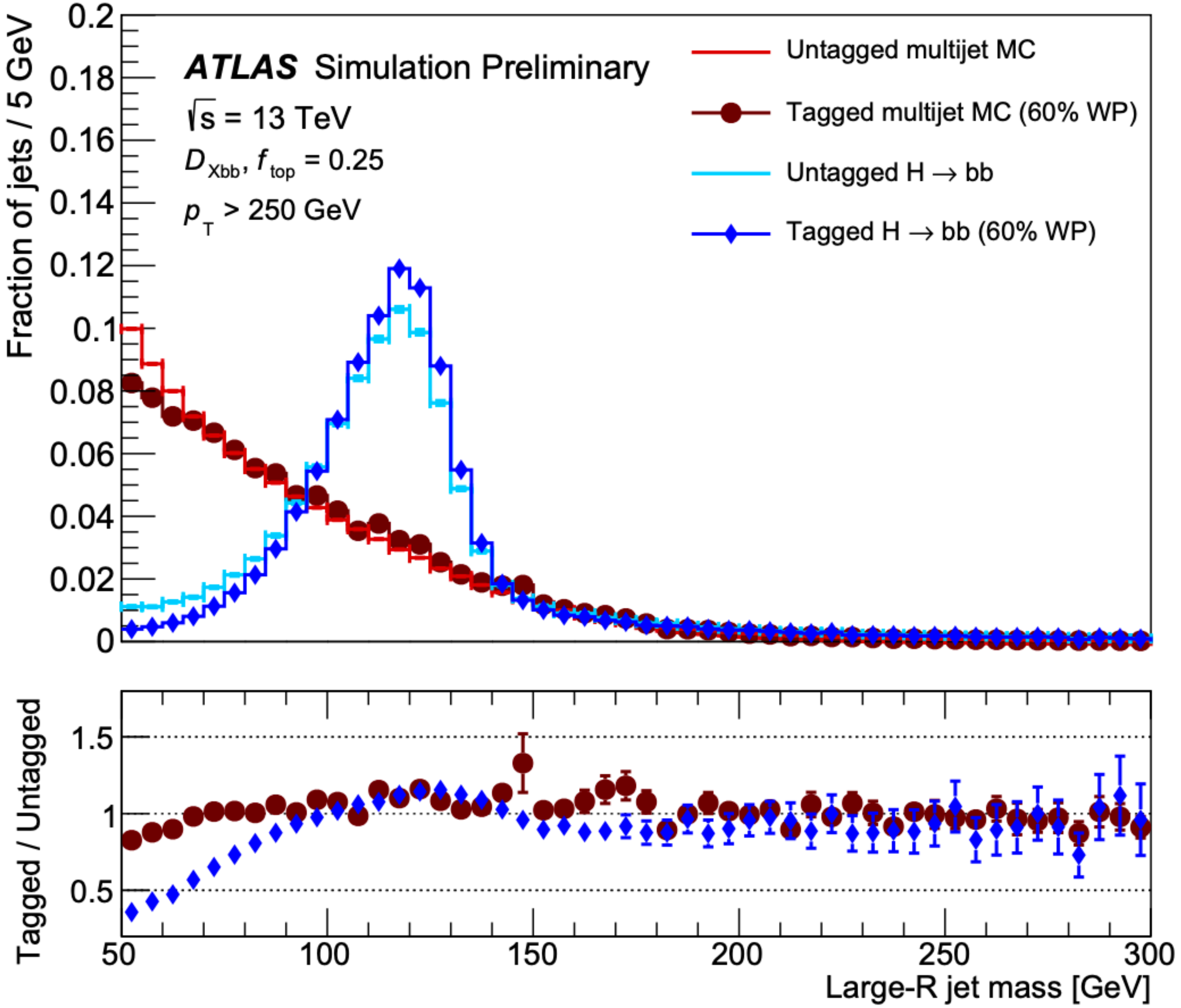
No mass sculpting of background!

No mass bias for searches!

Trained on resampled (pT, eta) distributions

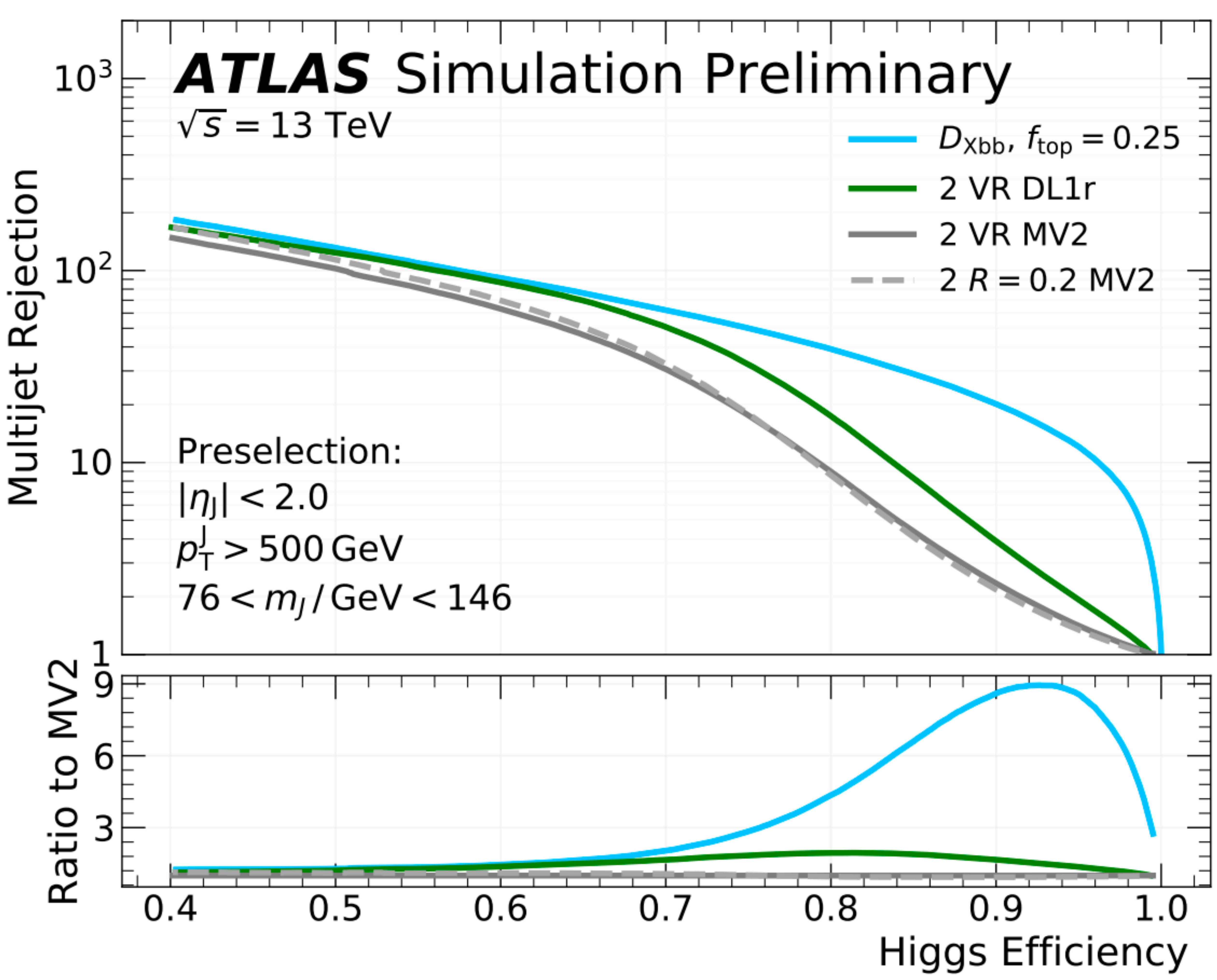


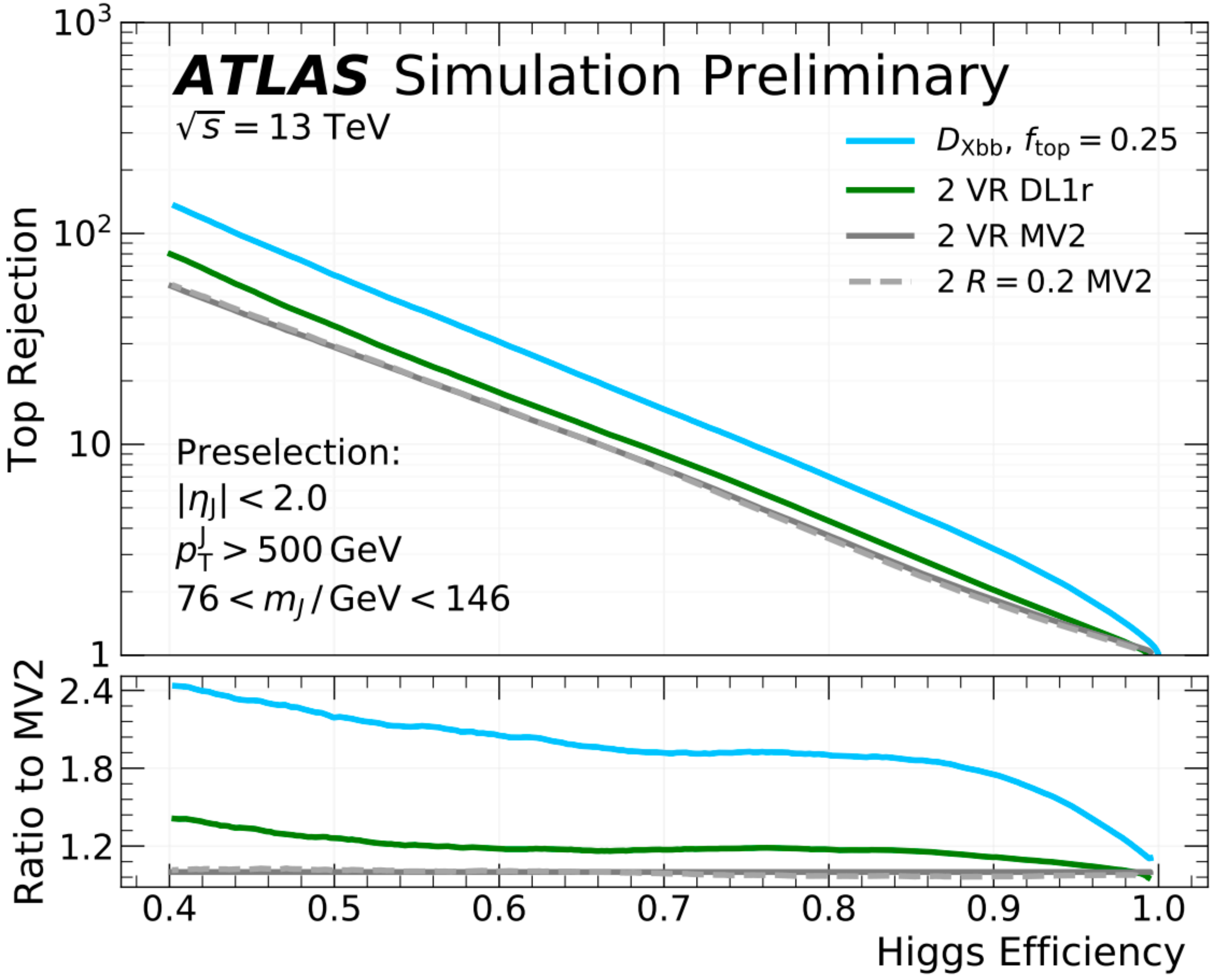
$$D_{Xbb} = \ln \frac{p_H}{f_{\text{top}} \cdot p_t + (1 - f_{\text{top}}) \cdot p_{\text{QCD}}}$$



Xbb Tagger: performance

Run-2 tagger





Tagged H(bb) jets / total H(bb) jets

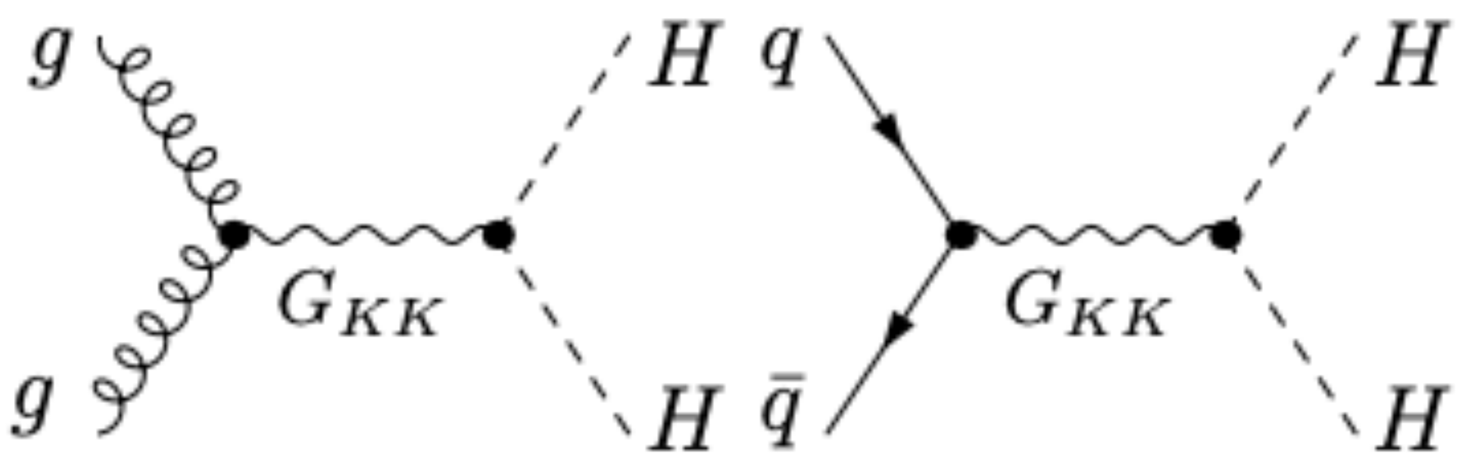
Xbb Tagger performance in G->H(bb)H(bb)

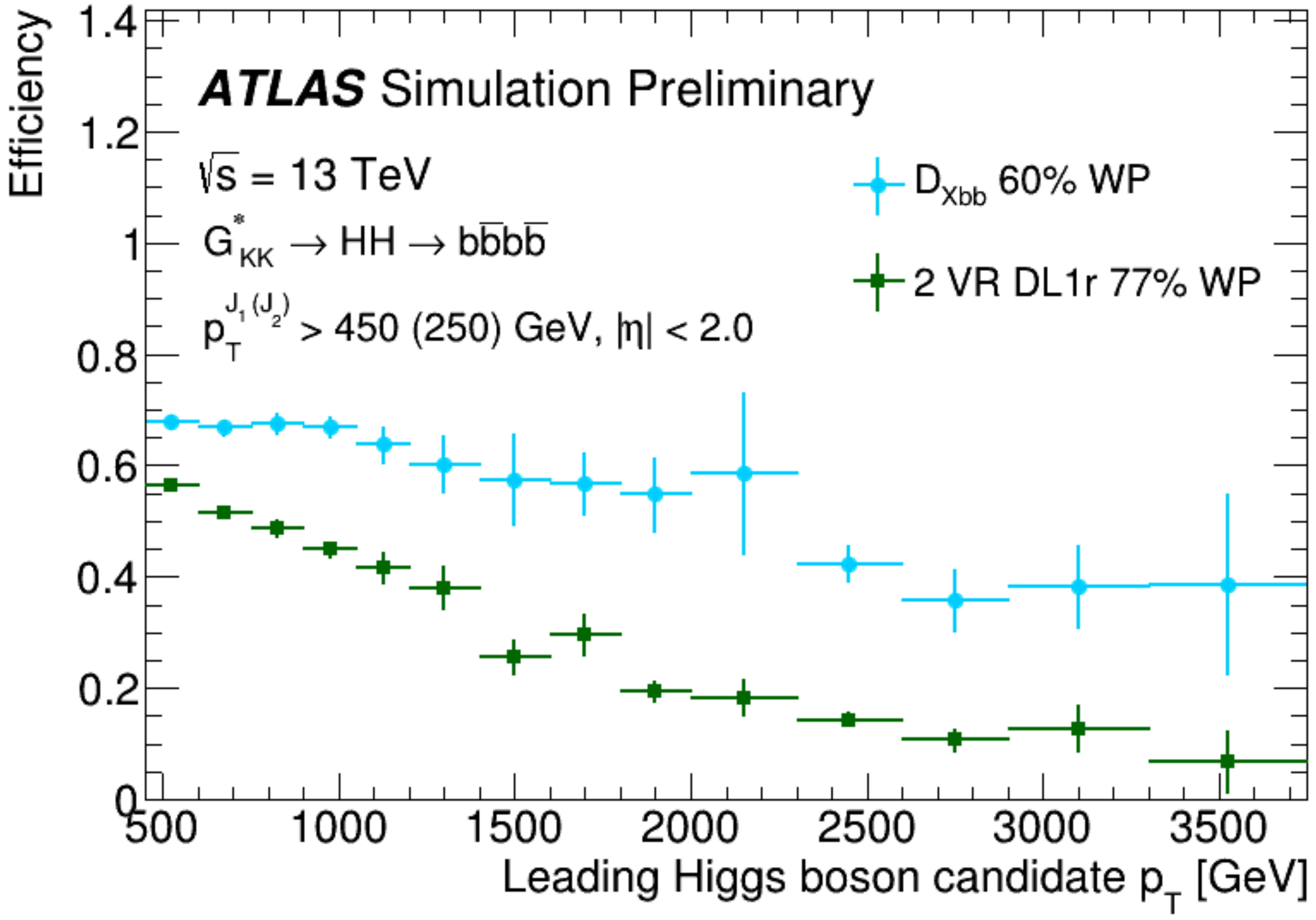
Run-2 tagger

Xbb was trained on signal samples of Randall-Sundrum Graviton
1 TeV < m(G) < 6TeV

Xbb is 20% to 110% more efficient than tagging two b-jets associated to Large-R jet in signal sample.

40% higher acceptance of background QCD multijet.

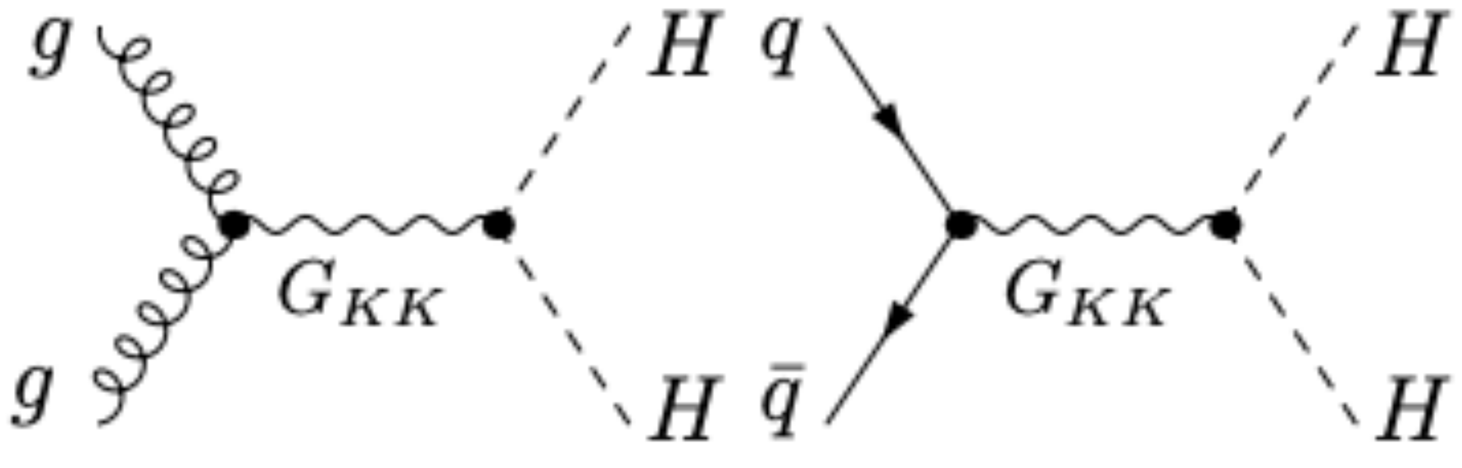


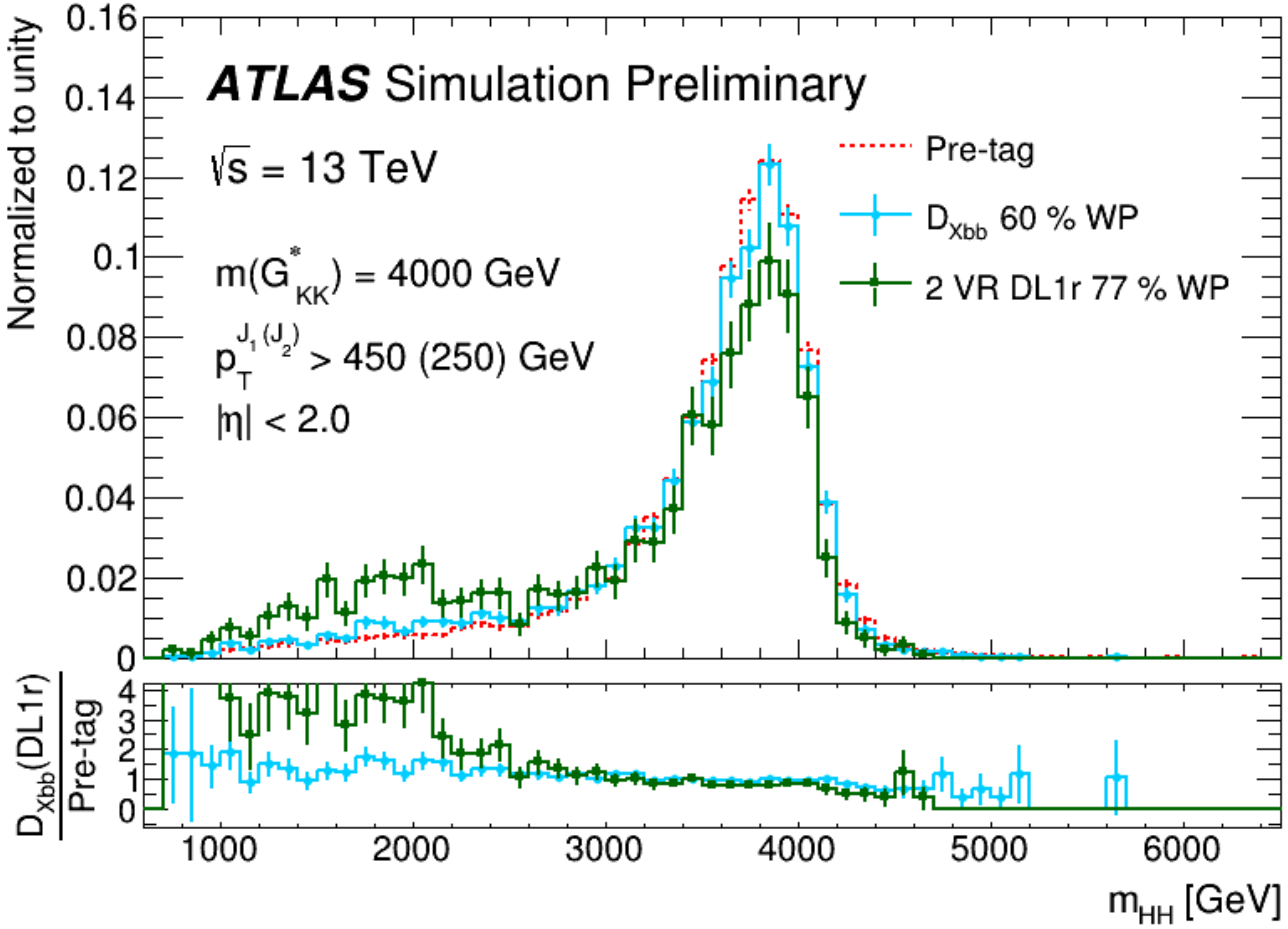


Xbb Tagger performance in G->H(bb)H(bb)

Run-2 tagger

Xbb reconstructed mHH after preselection fits pre-tag distribution better



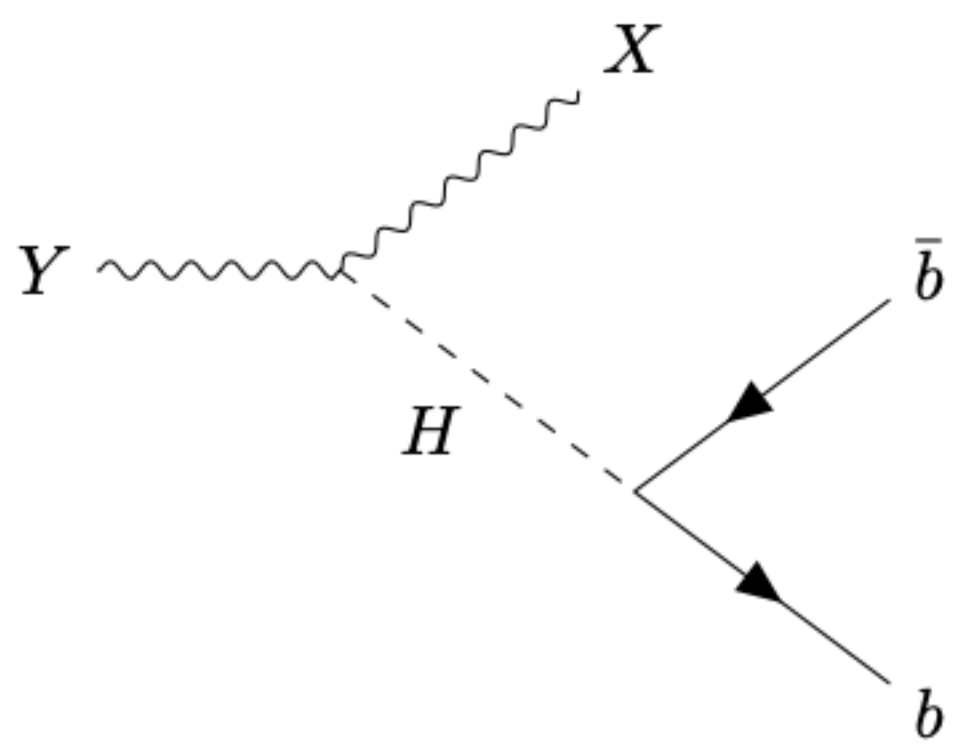


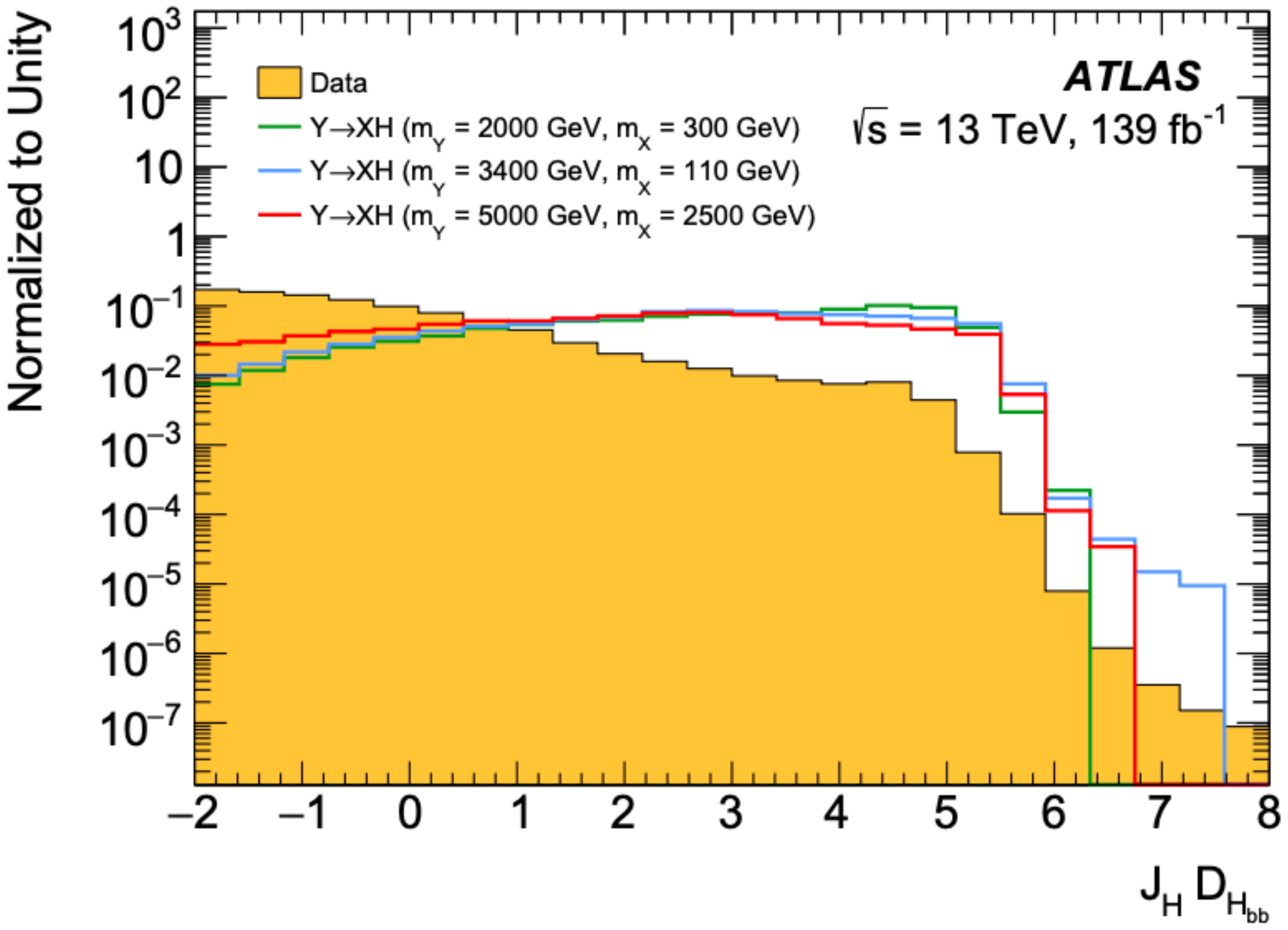
Xbb Tagger performance in Y->X(qq)H(bb)

Run-2 tagger

The Higgs Large-R Jet candidate is tagged using Xbb with 60% efficiency as the chosen working point.

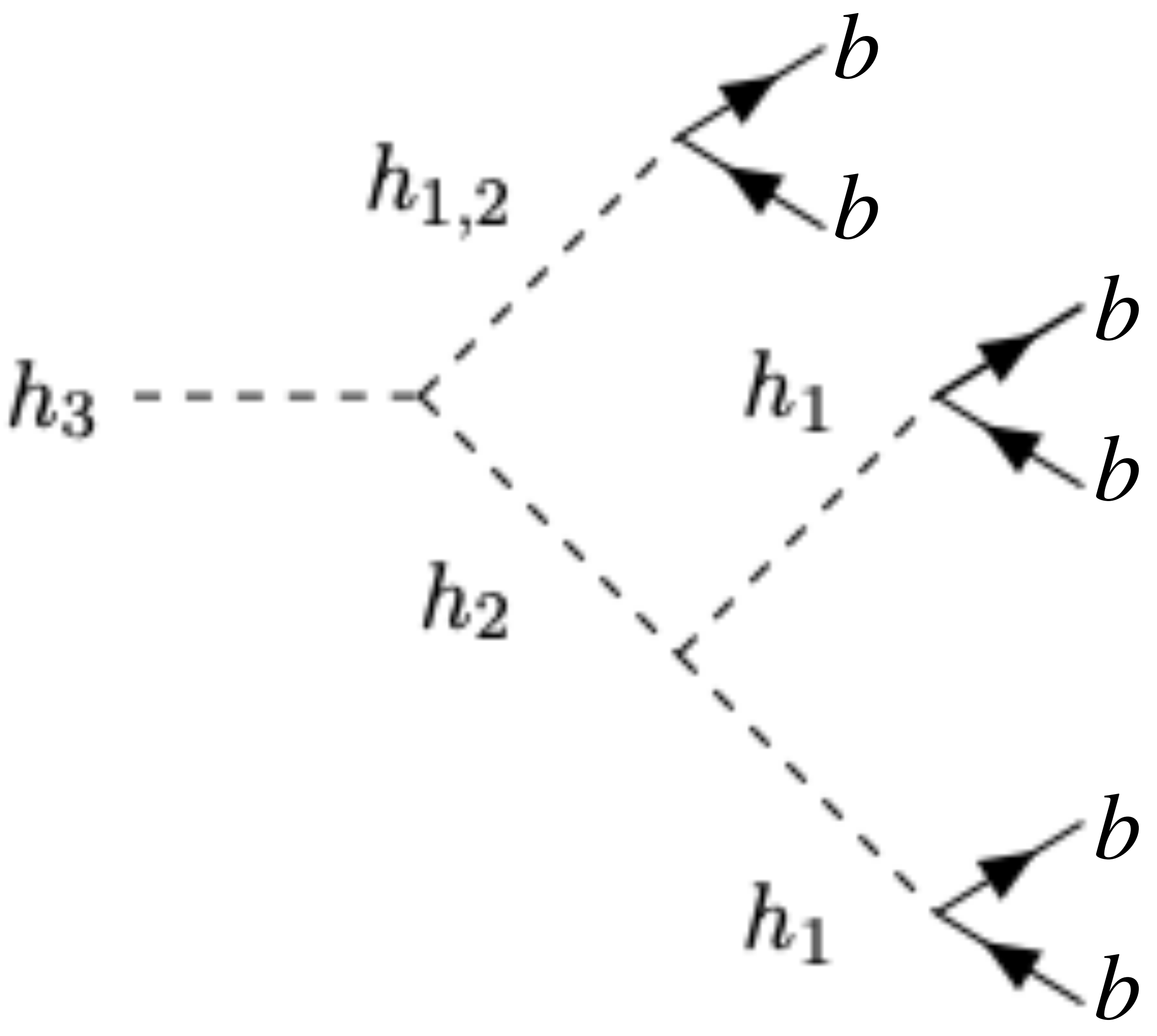
Best upper limit ~ 0.3 fb
(Recall 50 fb BSM HHH)





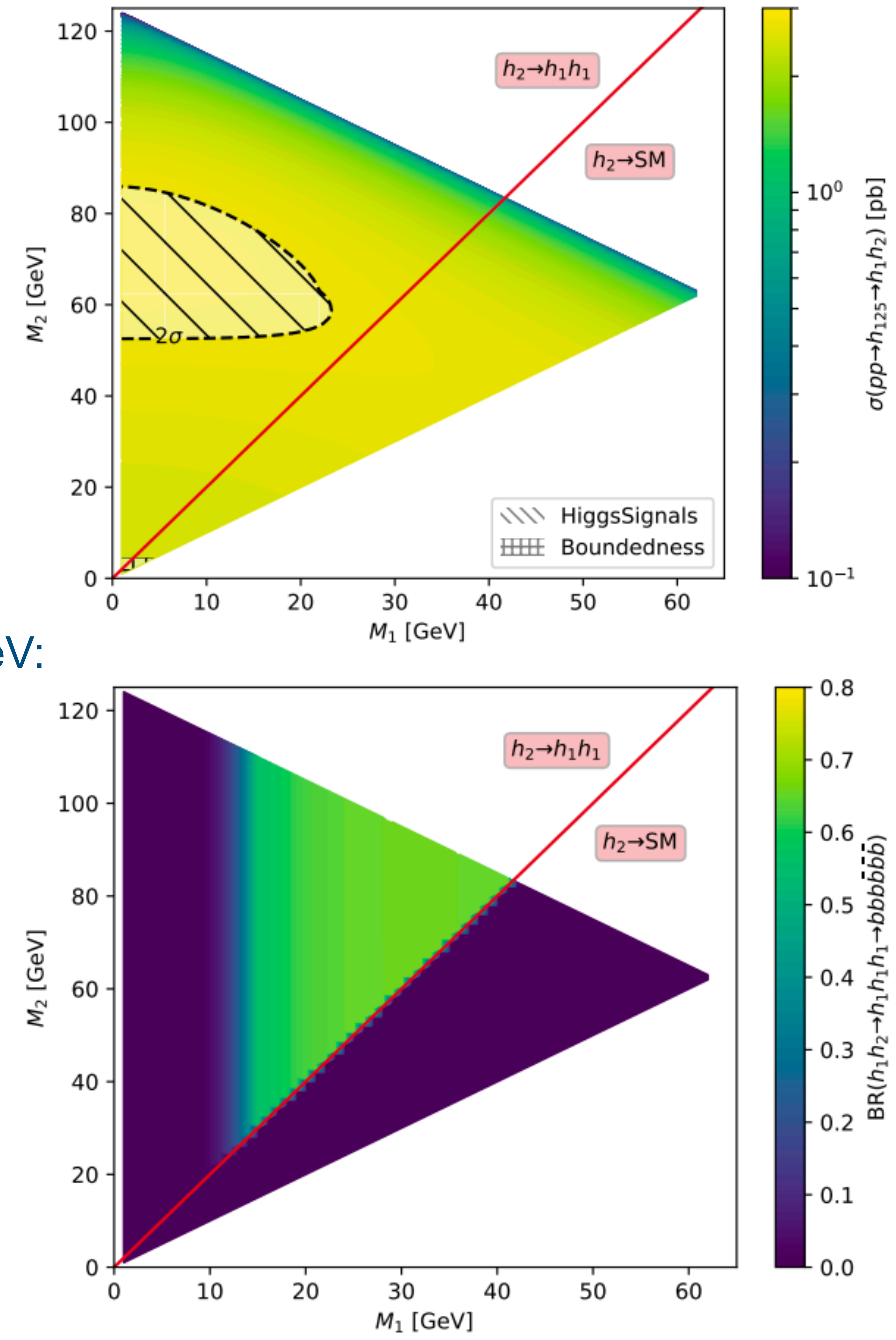
Lower mass resonances?

BSM models for HHH can have lower mass resonances
Still with 6b final state!



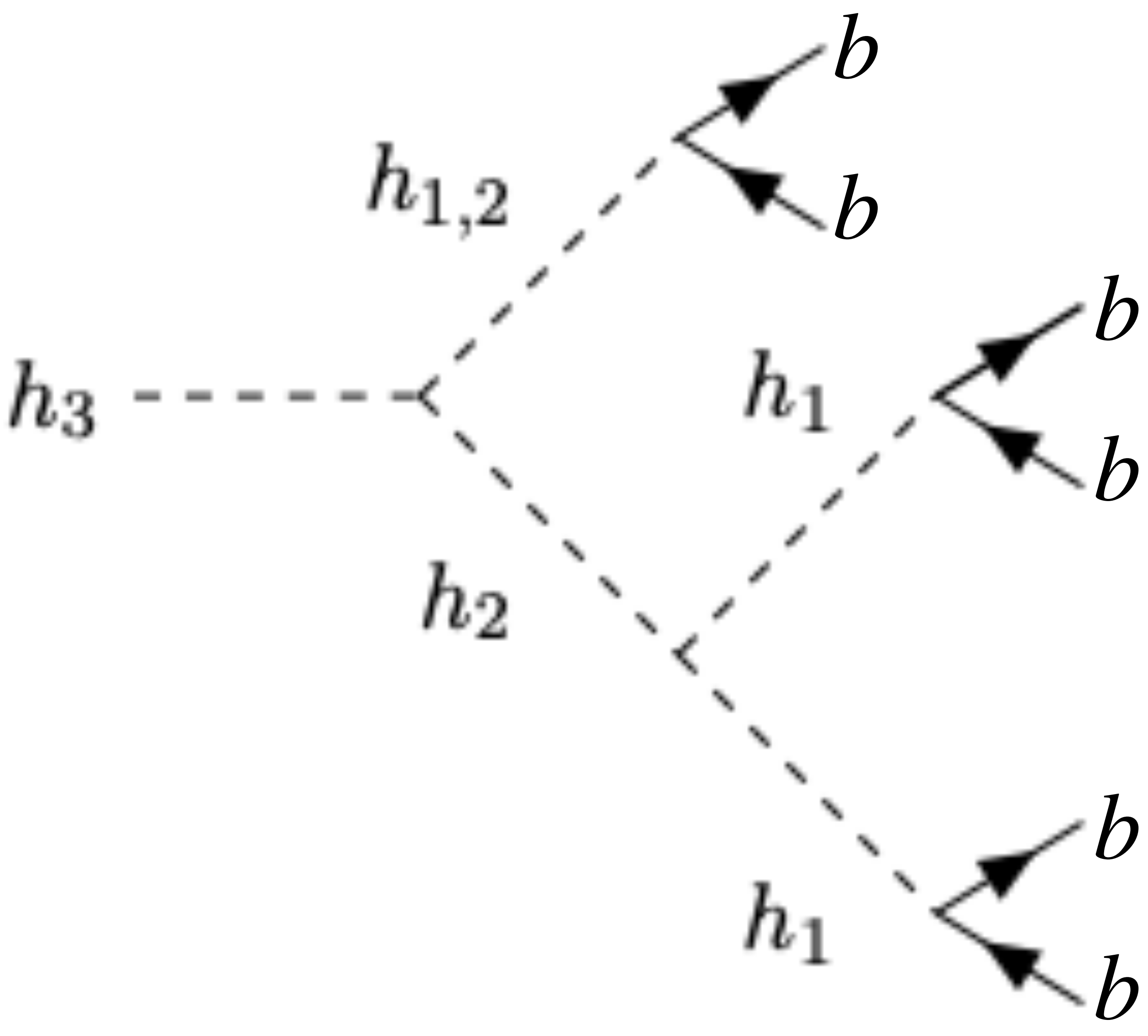
Our Higgs with m=125 GeV can be h3
h1 could have lower mass!

TRSM BP1 @13TeV:



Lower mass resonances?

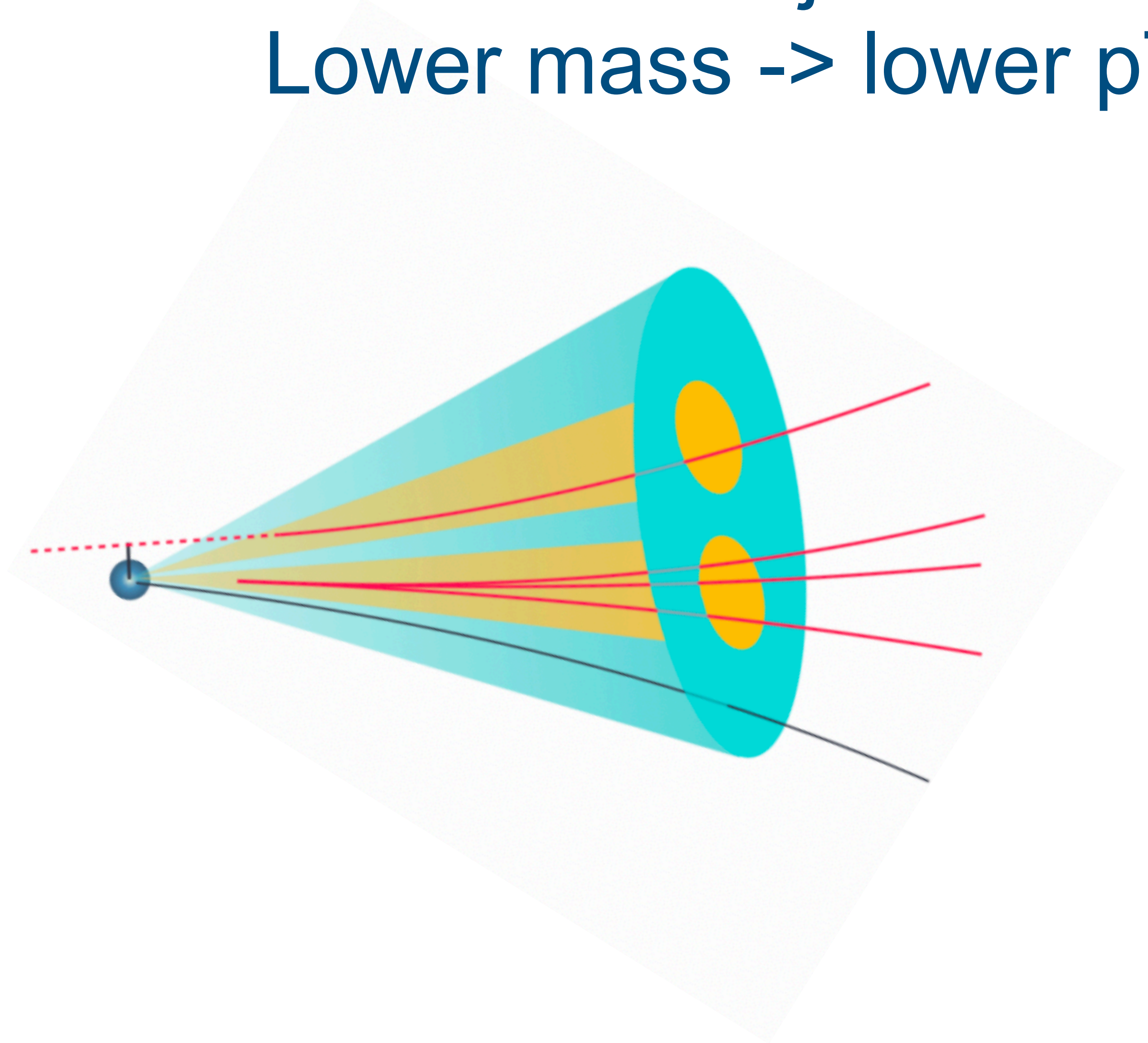
BSM models (scalars or pseudo scalars coupling to Higgs) for HHH can have lower mass resonances



Our Higgs with m=125 GeV can be h3
h1 could have lower mass!

$$R \sim 2m_h/pT$$

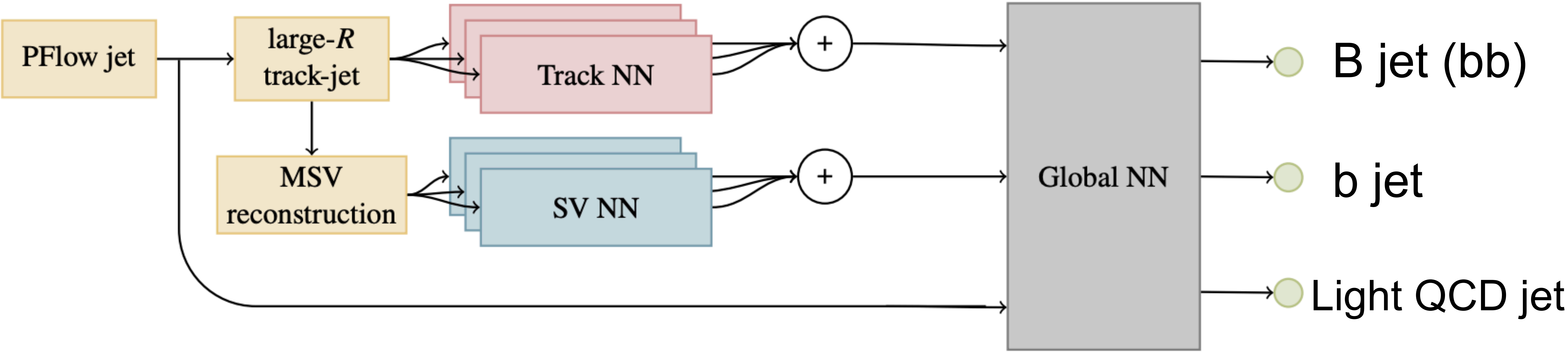
For the same jet radius,
Lower mass -> lower pT



DeXTer: low mass tagger

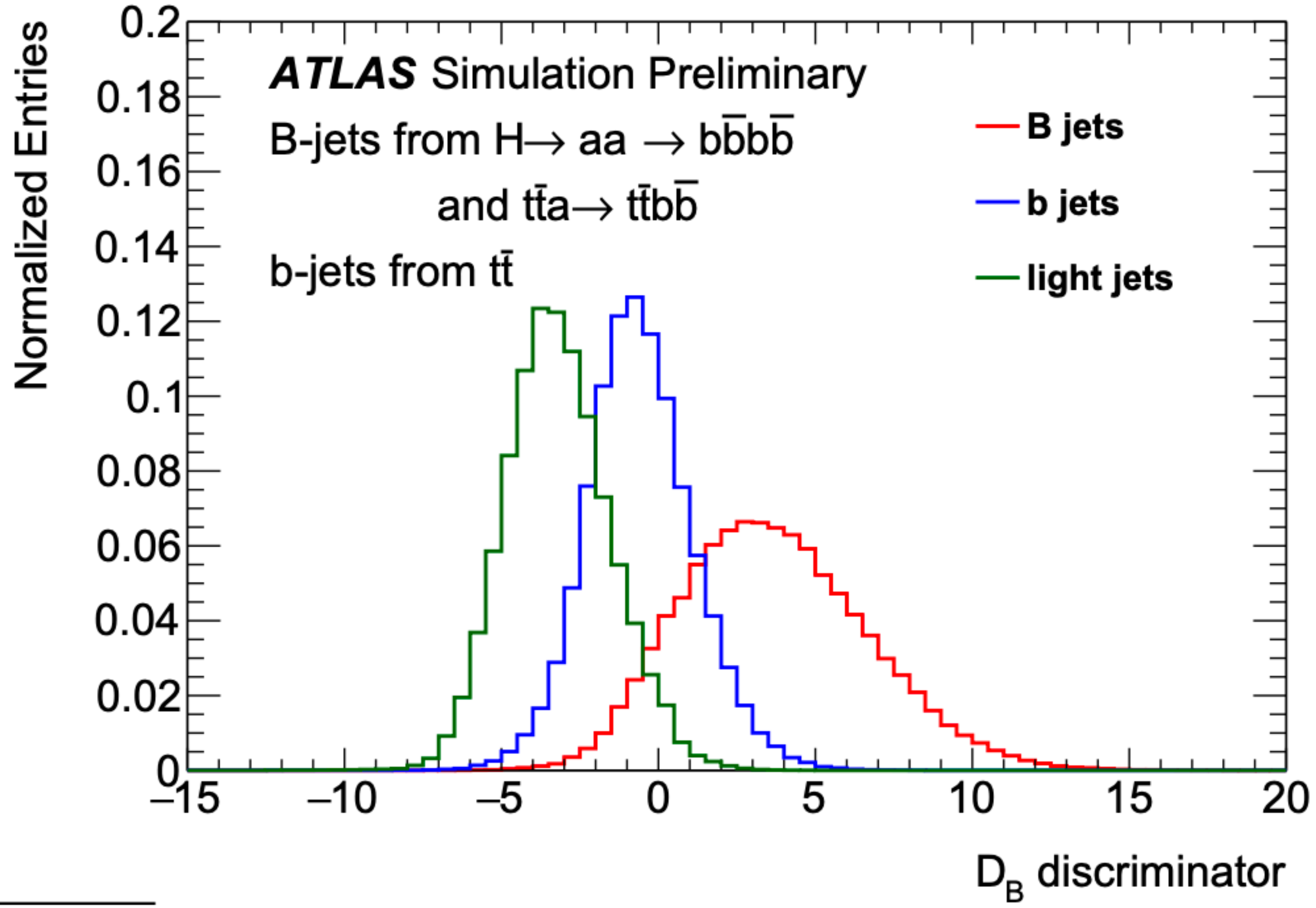
$R = 0.8$

DeXTer: Deep Sets based Neural Networks for Low-p_T $X \rightarrow b\bar{b}$ Identification in ATLAS



Trained on track information directly

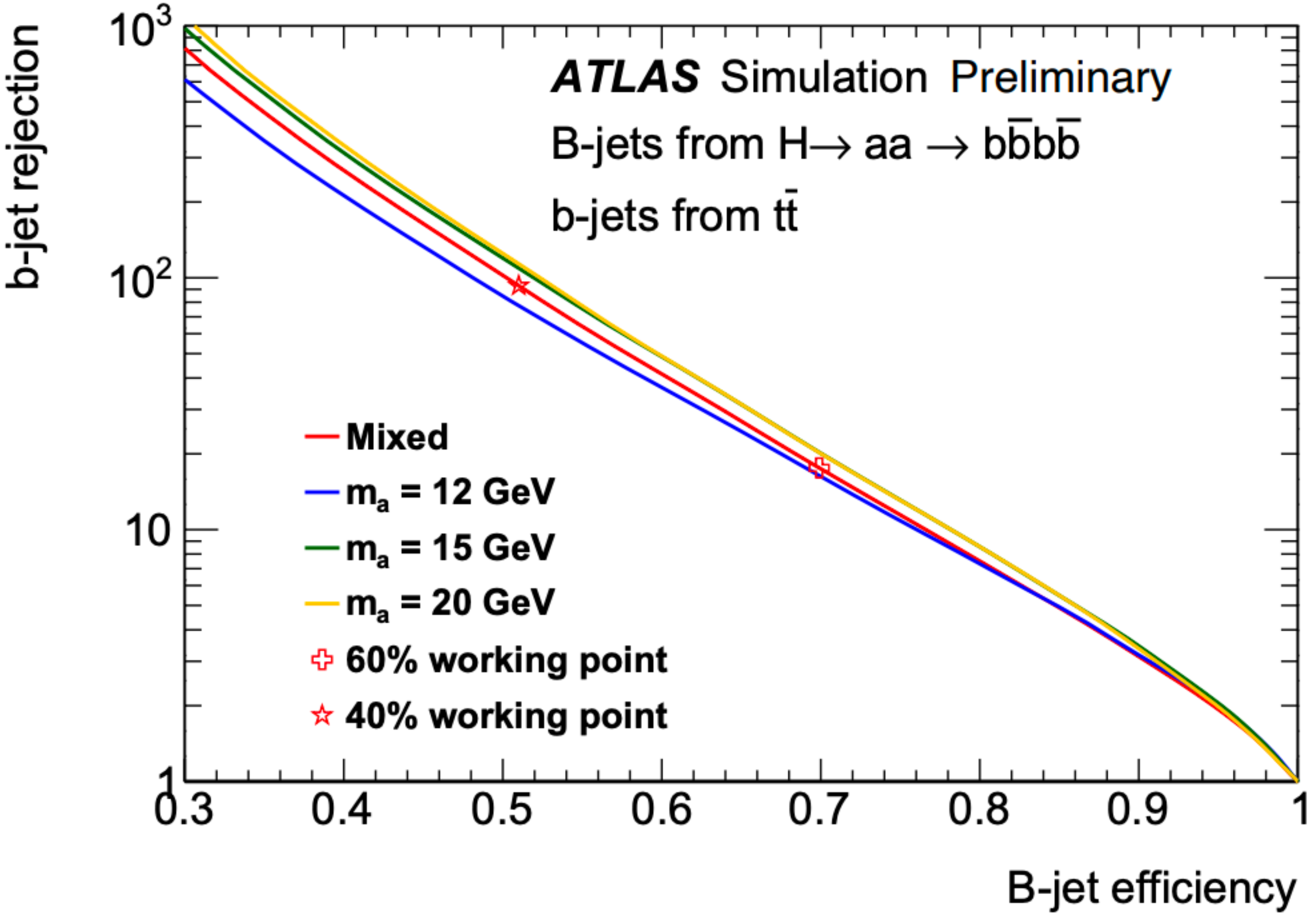
DeXTer: low mass tagger



B jet = bb jet

$$D_B = \ln \frac{p_B}{(1 - f_b)p_l + f_b p_b}$$

DeXTer: low mass tagger

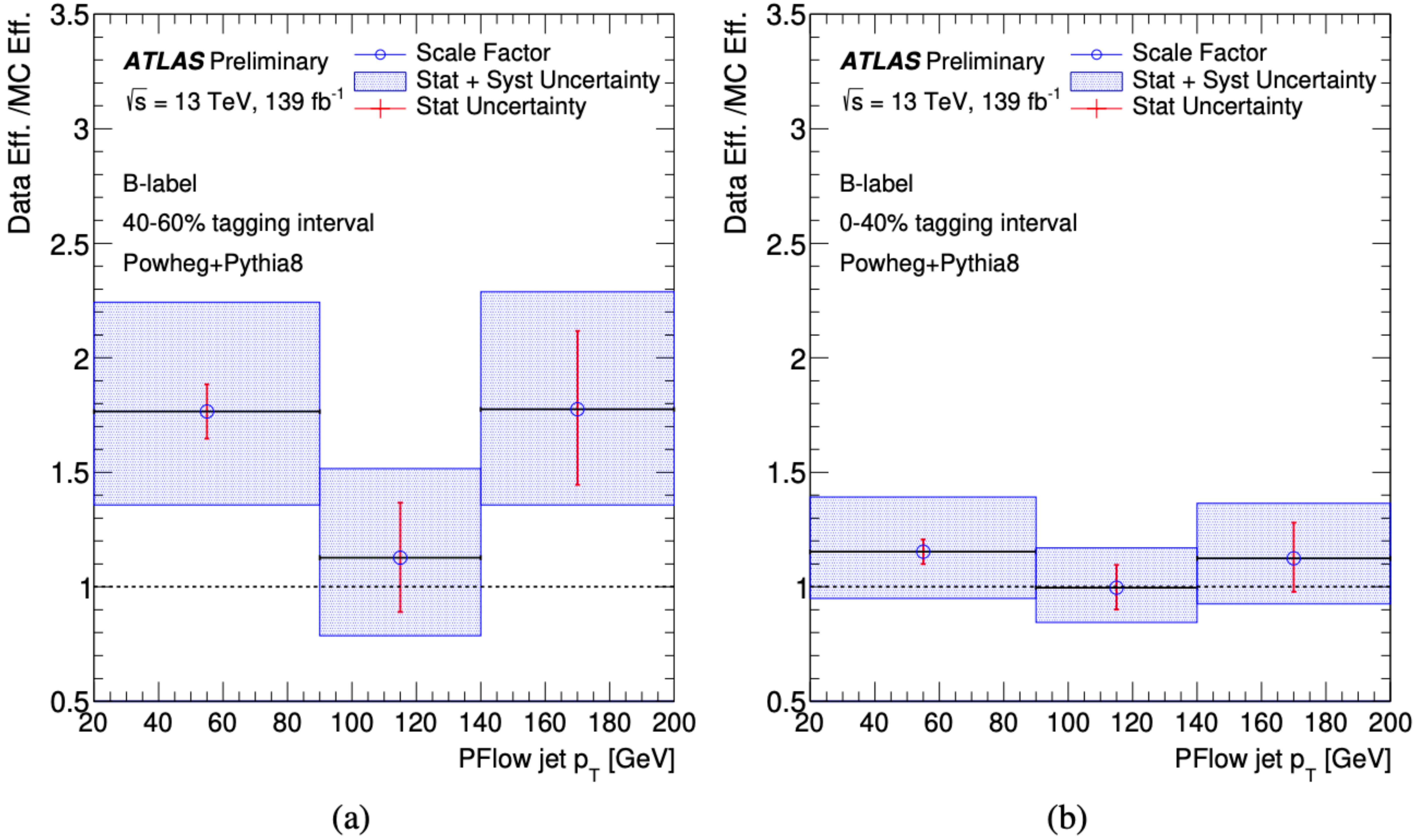


B jet = bb jet

Performance does not depend much on mass (nor colour)

DeXTer: low mass tagger: Calibration

using $Z(\rightarrow \ell^+\ell^-) + g(\rightarrow b\bar{b})$ events as a source of B-labeled jets and $t\bar{t}$ events events as a source of b-labeled jets

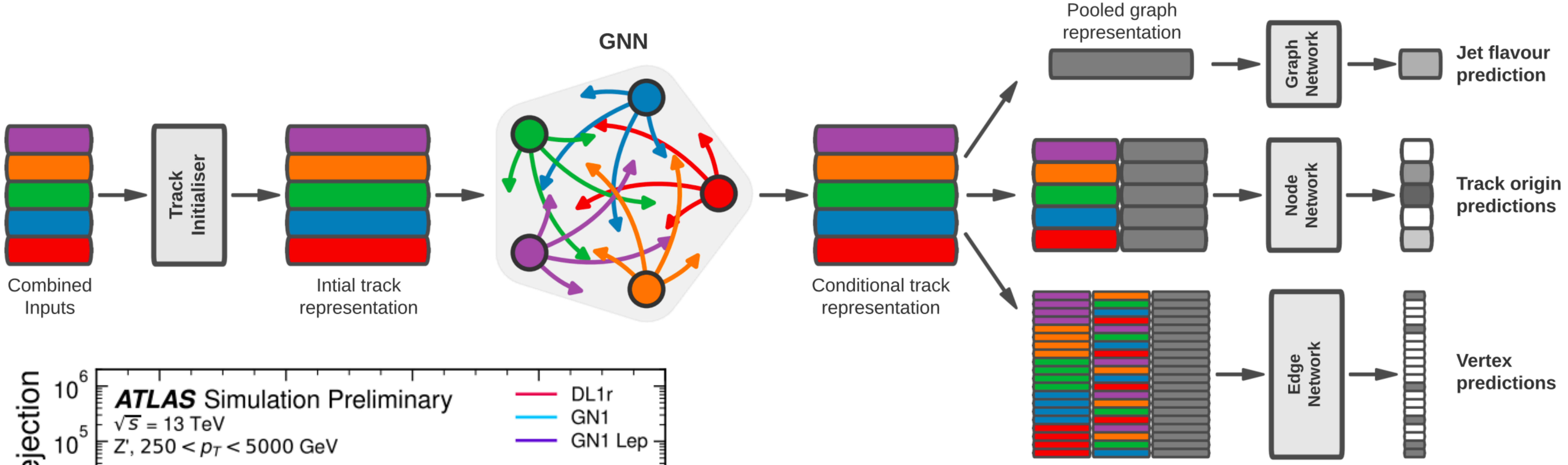


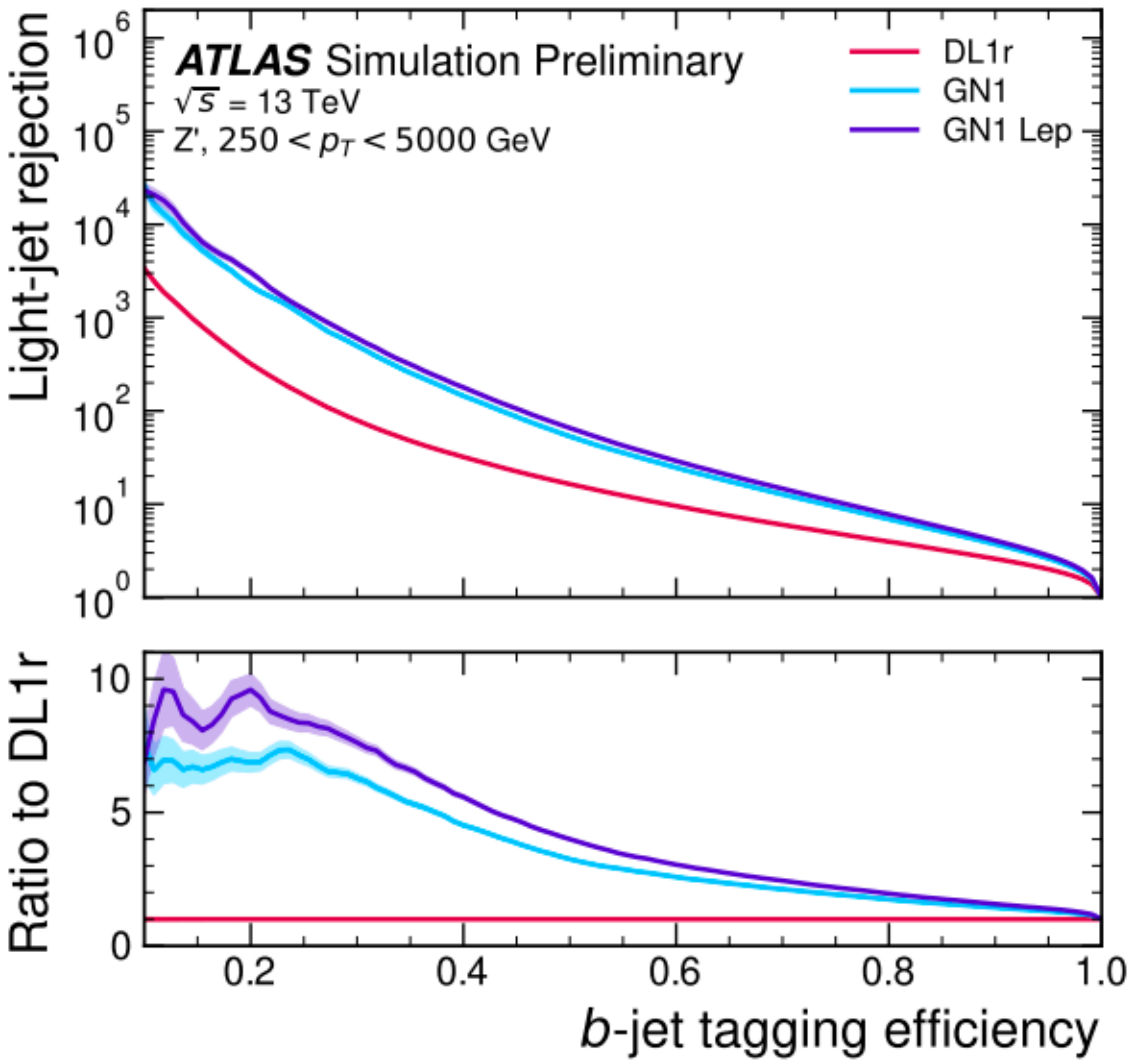
B jet = bb jet

Figure 8: Data-to-MC scale factors obtained for B-labeled jets in the (a) 40 – 60% and (b) 0 – 40% tagging intervals. The blue error band includes systematic and statistical uncertainties. The red error bar represents the statistical uncertainty only.

GNNs: Future of X->bb tagging

For Run-3





Improved performance in single-b tagging!

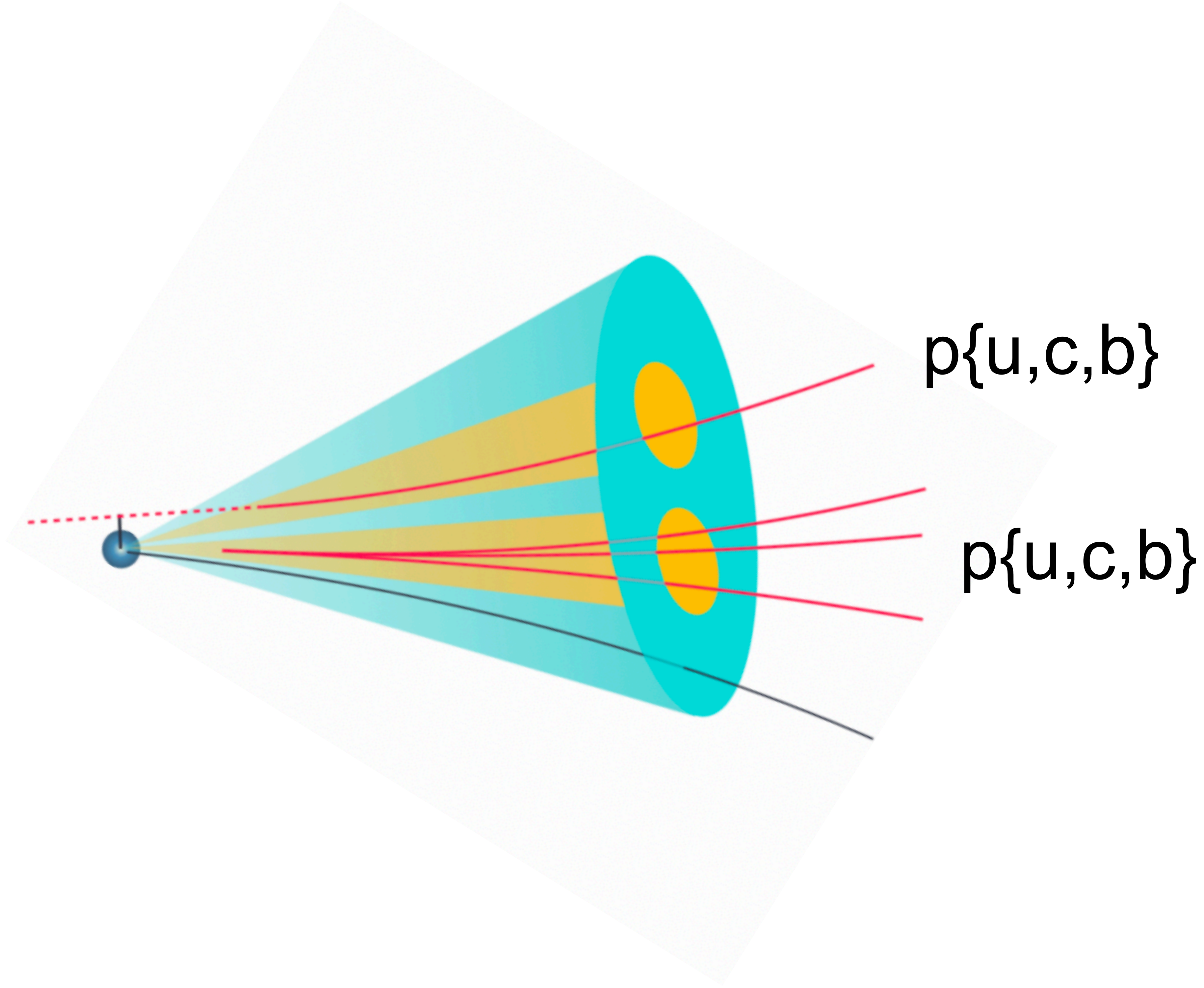
Since GN1 does not have any assumptions on the jet collection, it treats Large-R and small-R jets the same.

->Expect improvements in Xbb tagging as well!

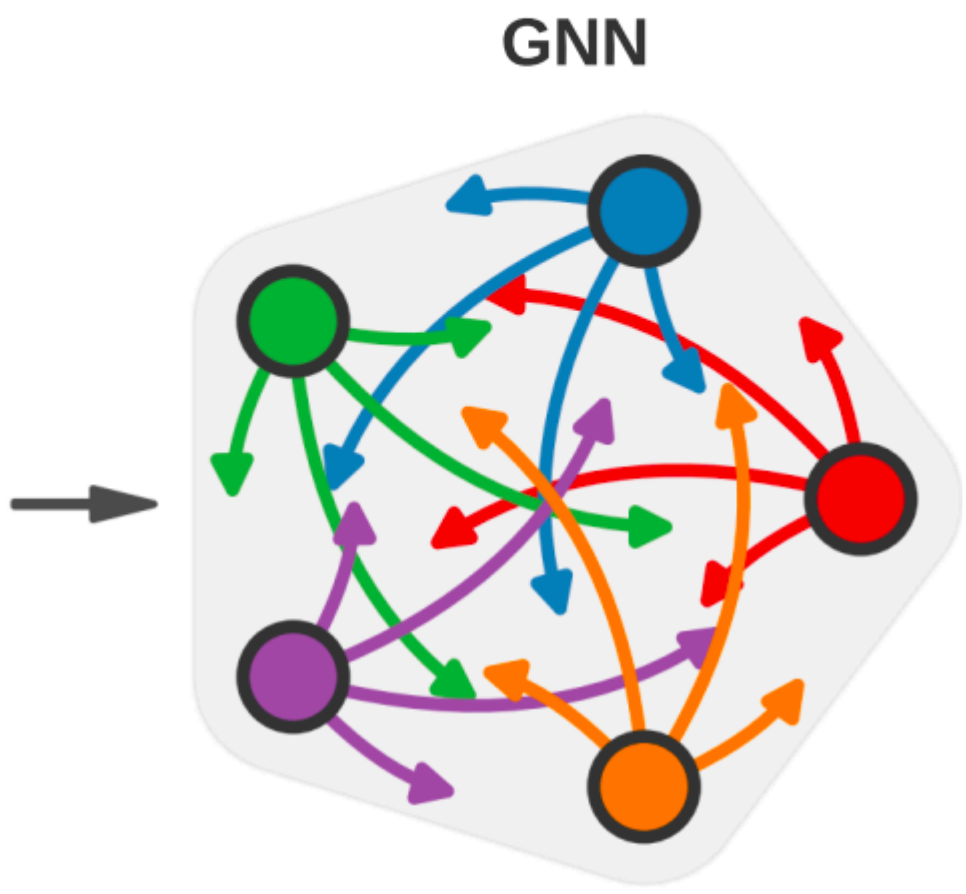
Boosted bb tagging is useful!

References: (also if you click on figures)
https://cds.cern.ch/record/2724739/files/ATL-PHYS-PUB-2020-019.pdf
https://arxiv.org/pdf/2211.16345.pdf
https://arxiv.org/pdf/1906.11005.pdf
https://arxiv.org/pdf/2306.03637.pdf
https://arxiv.org/pdf/1703.10485.pdf
http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2022-002/

Boosted bb tagging proved useful for SM HH
Run-3 boosted SM HH analyses coming…
Boosted HHH searches could be useful as well



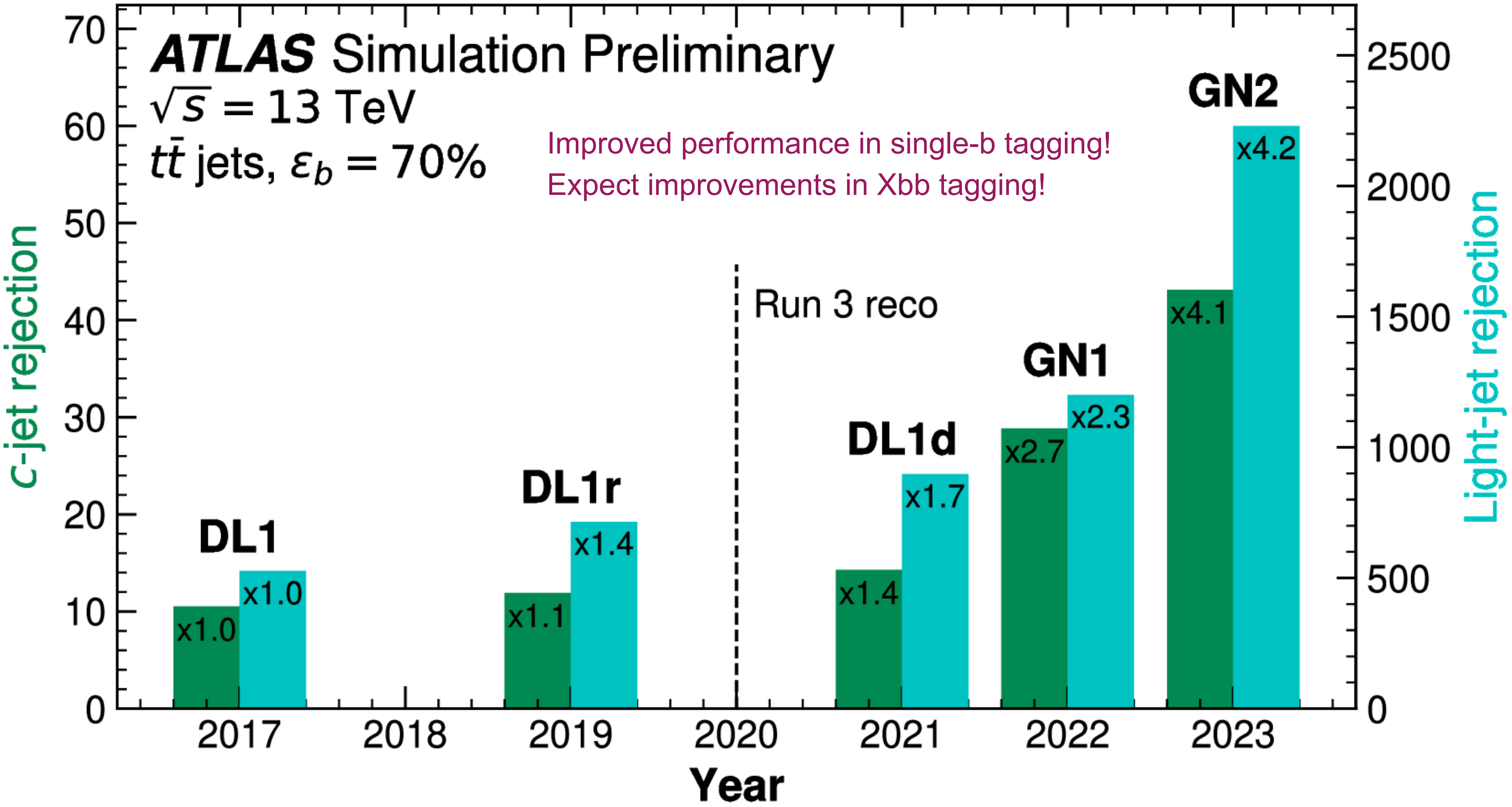
dR = 1.0



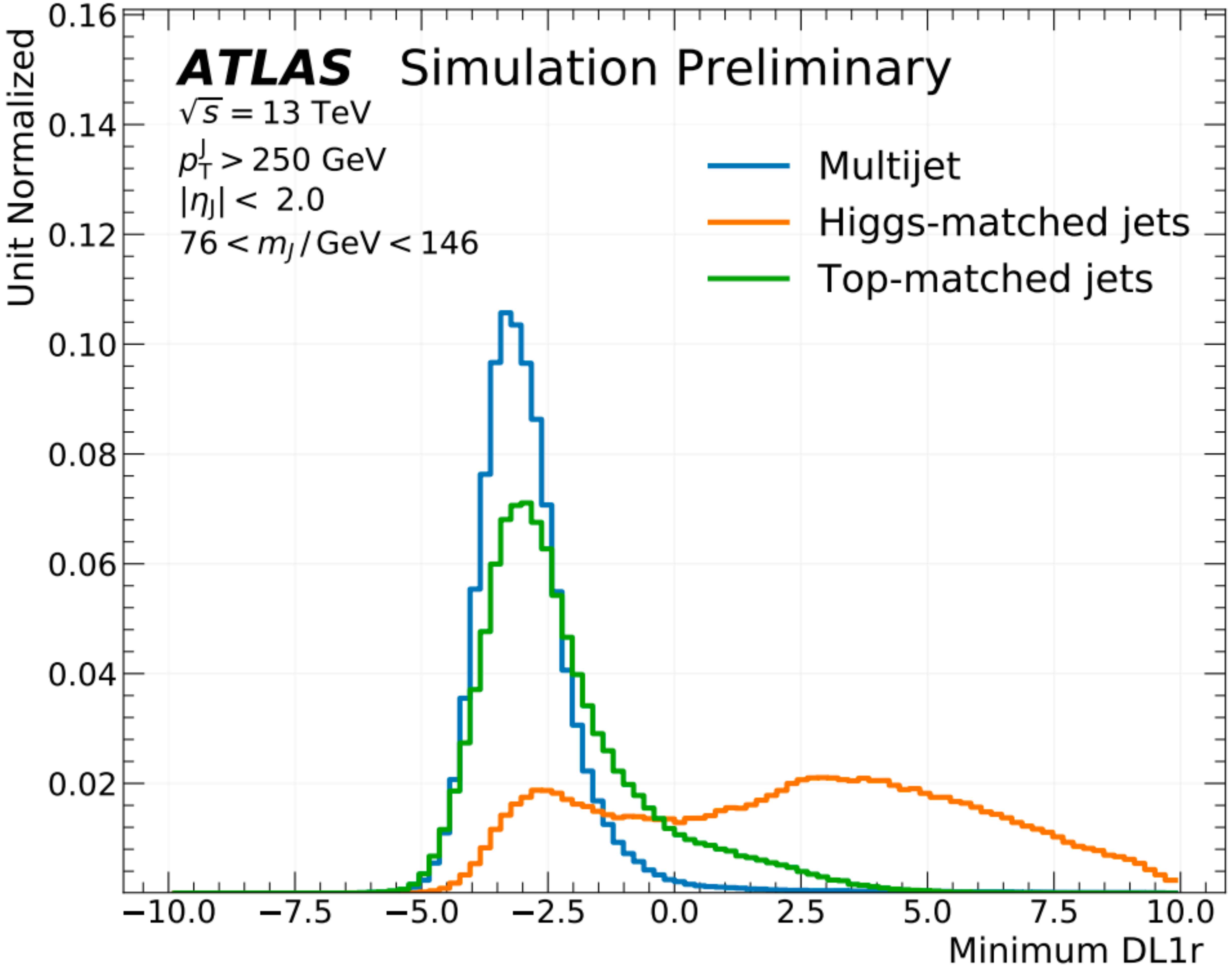
Back up

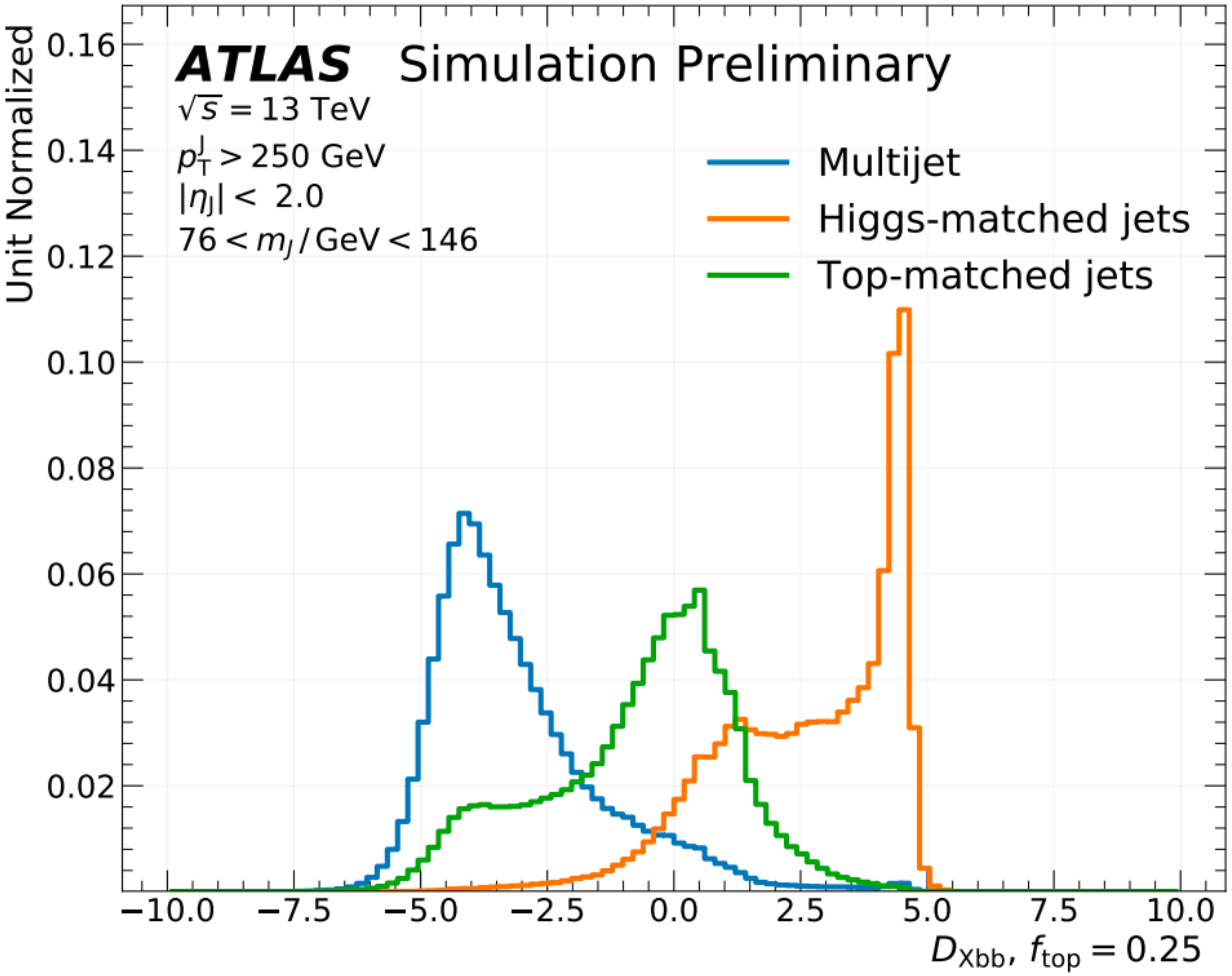
GNNs: Future of X->bb tagging

For Run-3

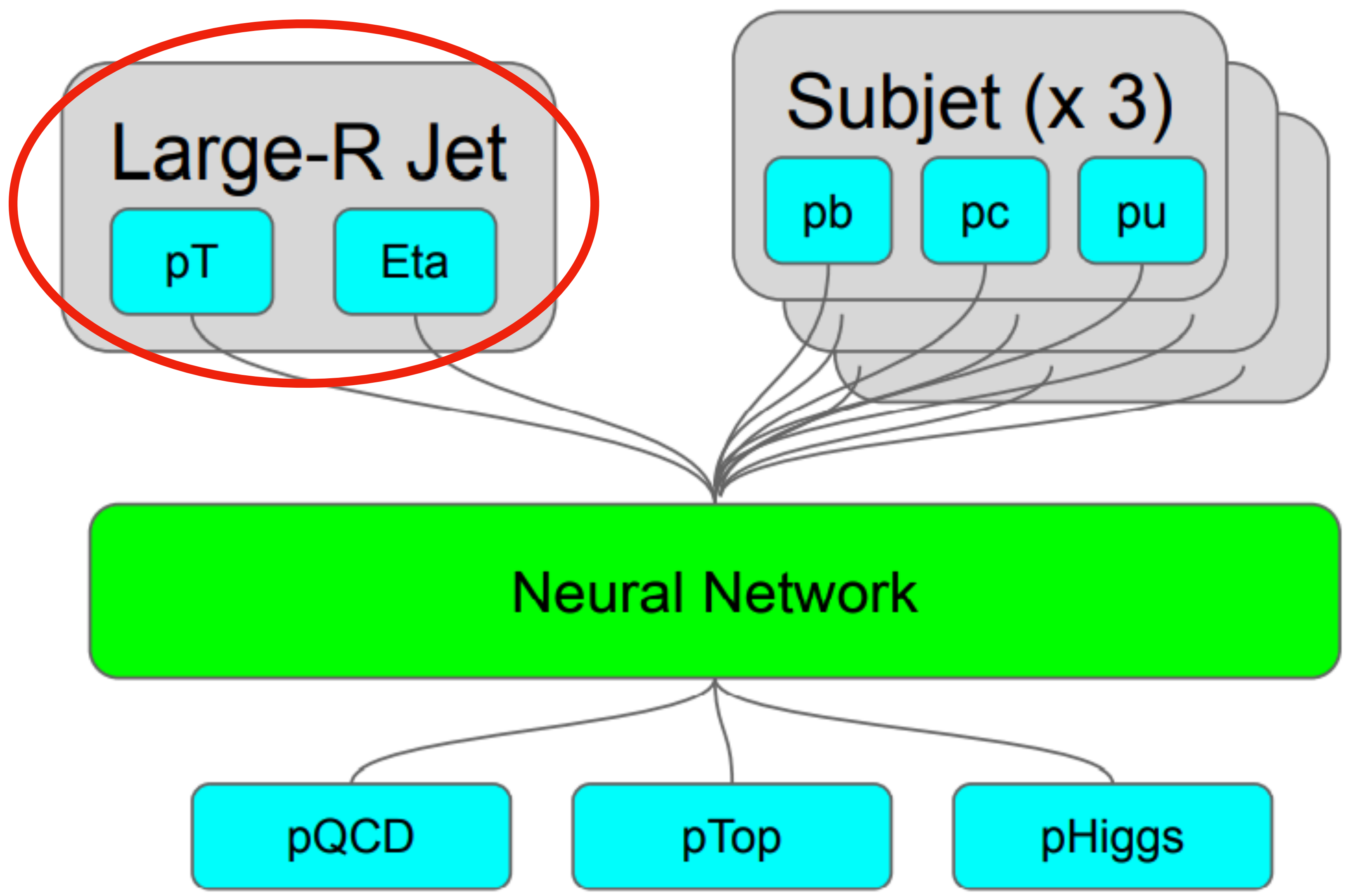


Back up


ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV
$p_T^J > 250$ GeV
$|\eta_J| < 2.0$
$76 < m_J\,/\,\mathrm{GeV} < 146$
Multijet
Higgs-matched jets
Top-matched jets
Unit Normalized
Minimum DL1r


ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV
$p_T^J > 250$ GeV
$|\eta_J| < 2.0$
$76 < m_J\,/\,\mathrm{GeV} < 146$
Multijet
Higgs-matched jets
Top-matched jets
Unit Normalized
D_{Xbb}, $f_{top} = 0.25$

Xbb Tagger: resampling



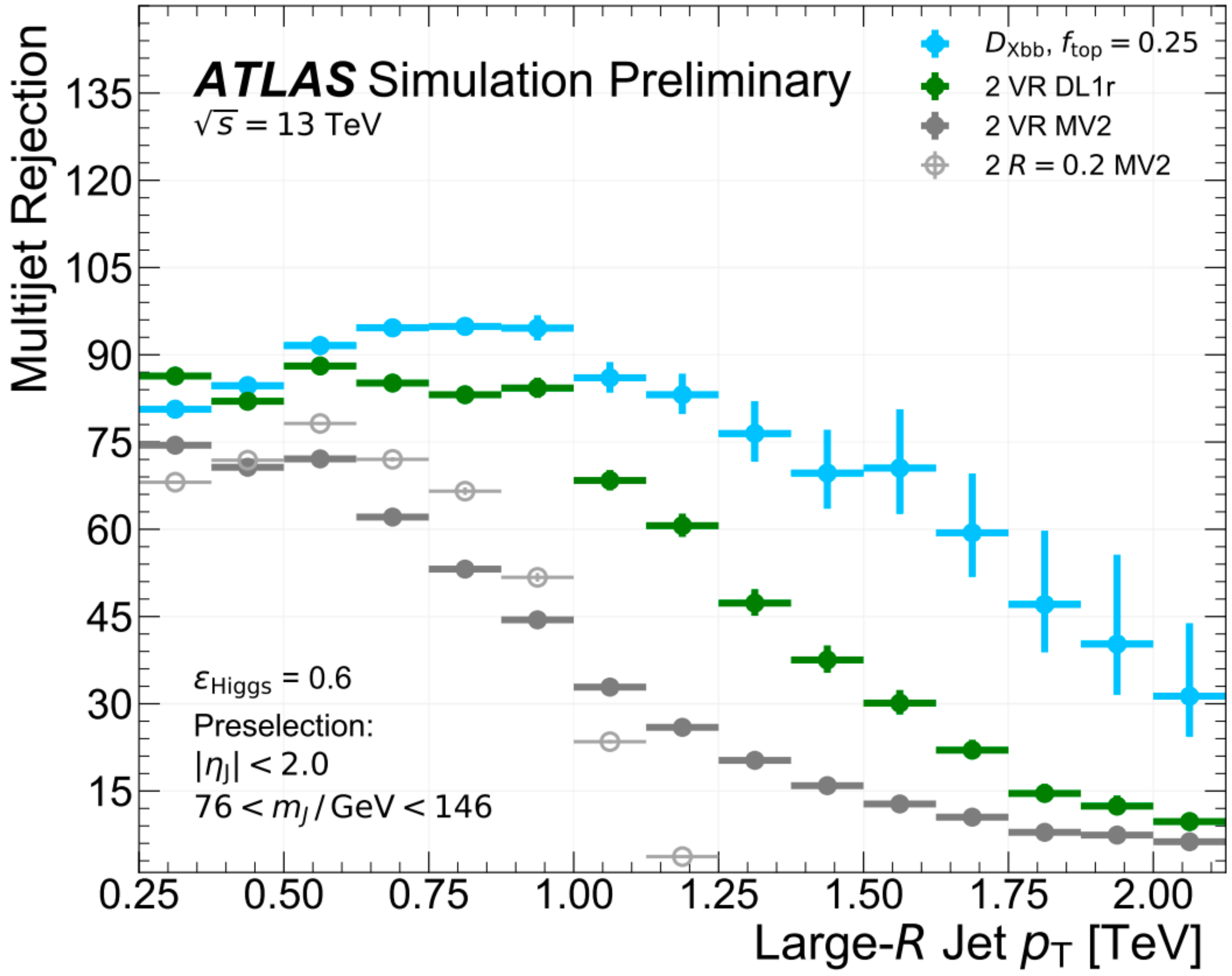
Before training!

Resampling: We remove from our samples such that (pT, eta) distributions of signal and backgrounds are the same!

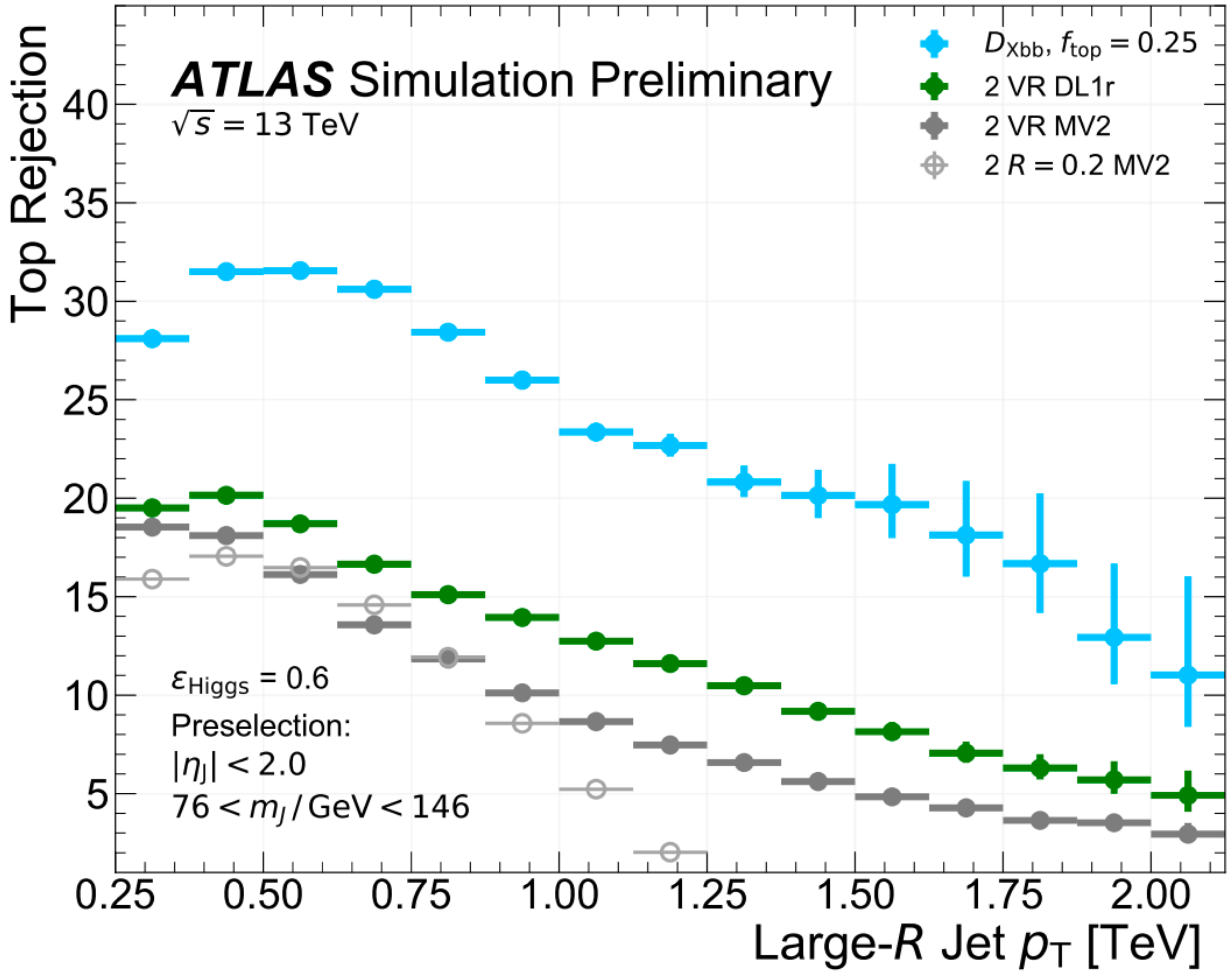
(pT, eta) correlate with the mass of the large-R jet.
We want the tagger to be mass-agnostic.

PLOTS

Back up



ATLAS Simulation Preliminary
$\sqrt{s} = 13$ TeV
D_{Xbb}, $f_{top} = 0.25$
2 VR DL1r
2 VR MV2
2 $R = 0.2$ MV2
$\varepsilon_{Higgs} = 0.6$
Preselection:
$|\eta_J| < 2.0$
$76 < m_J$ / GeV < 146
Multijet Rejection
Large-R Jet p_T [TeV]

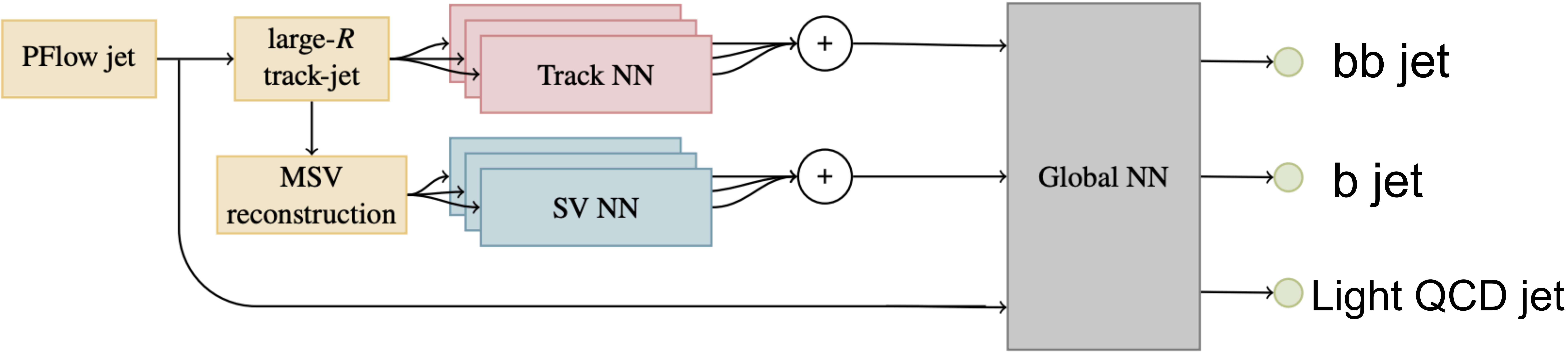


ATLAS Simulation Preliminary
$\sqrt{s} = 13$ TeV
D_{Xbb}, $f_{top} = 0.25$
2 VR DL1r
2 VR MV2
2 $R = 0.2$ MV2
$\varepsilon_{Higgs} = 0.6$
Preselection:
$|\eta_J| < 2.0$
$76 < m_J$ / GeV < 146
Top Rejection
Large-R Jet p_T [TeV]

DeXTer: for low mass

$R = 0.8$

DeXTer: Deep Sets based Neural Networks for Low-p_T $X \to b\bar{b}$ Identification in ATLAS

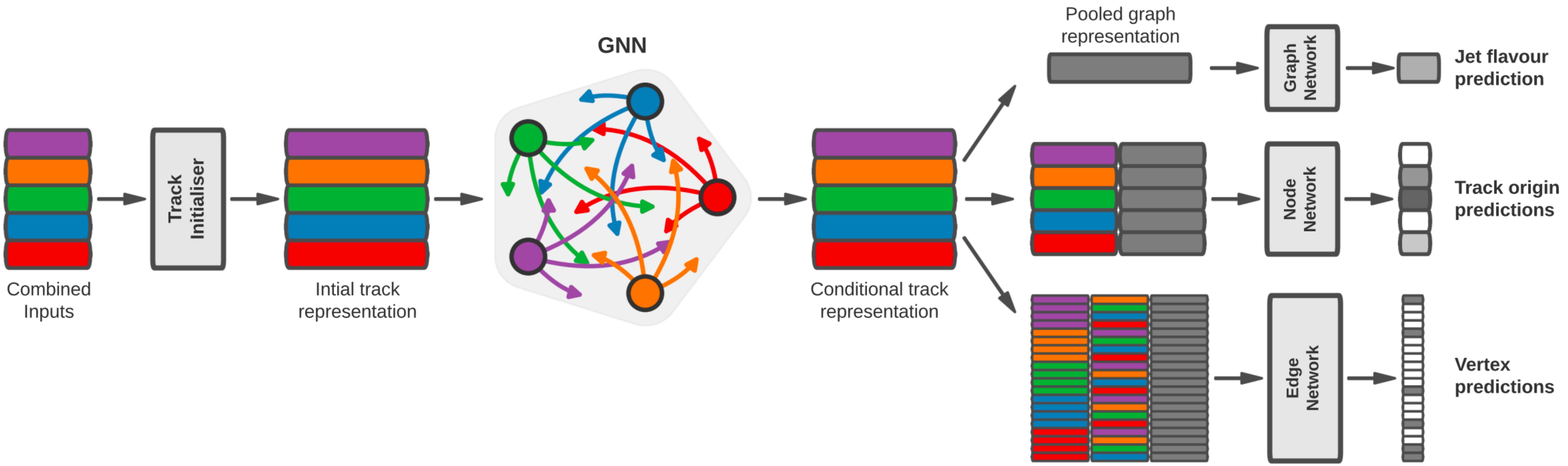


Permutation invariant Deep Set model

large-R track jet: Pflow jet of radius 0.4 reclustered with tracks nearby
+Multiple secondary vertices + kinematics of the PFlow Jet (pT and eta)

GNNs: Future of X->bb tagging

For Run-3

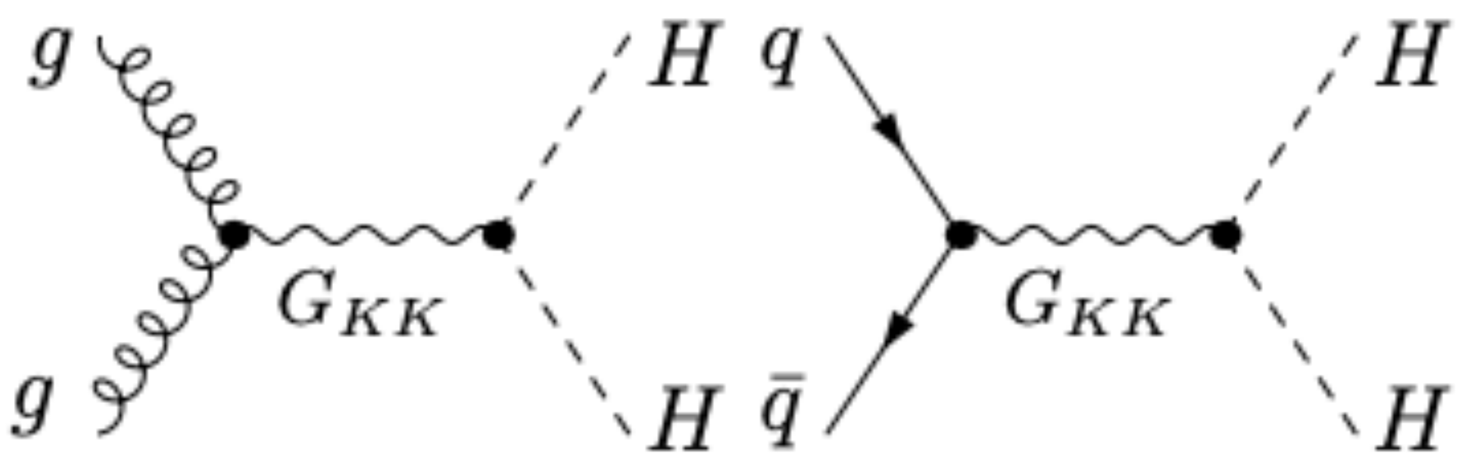


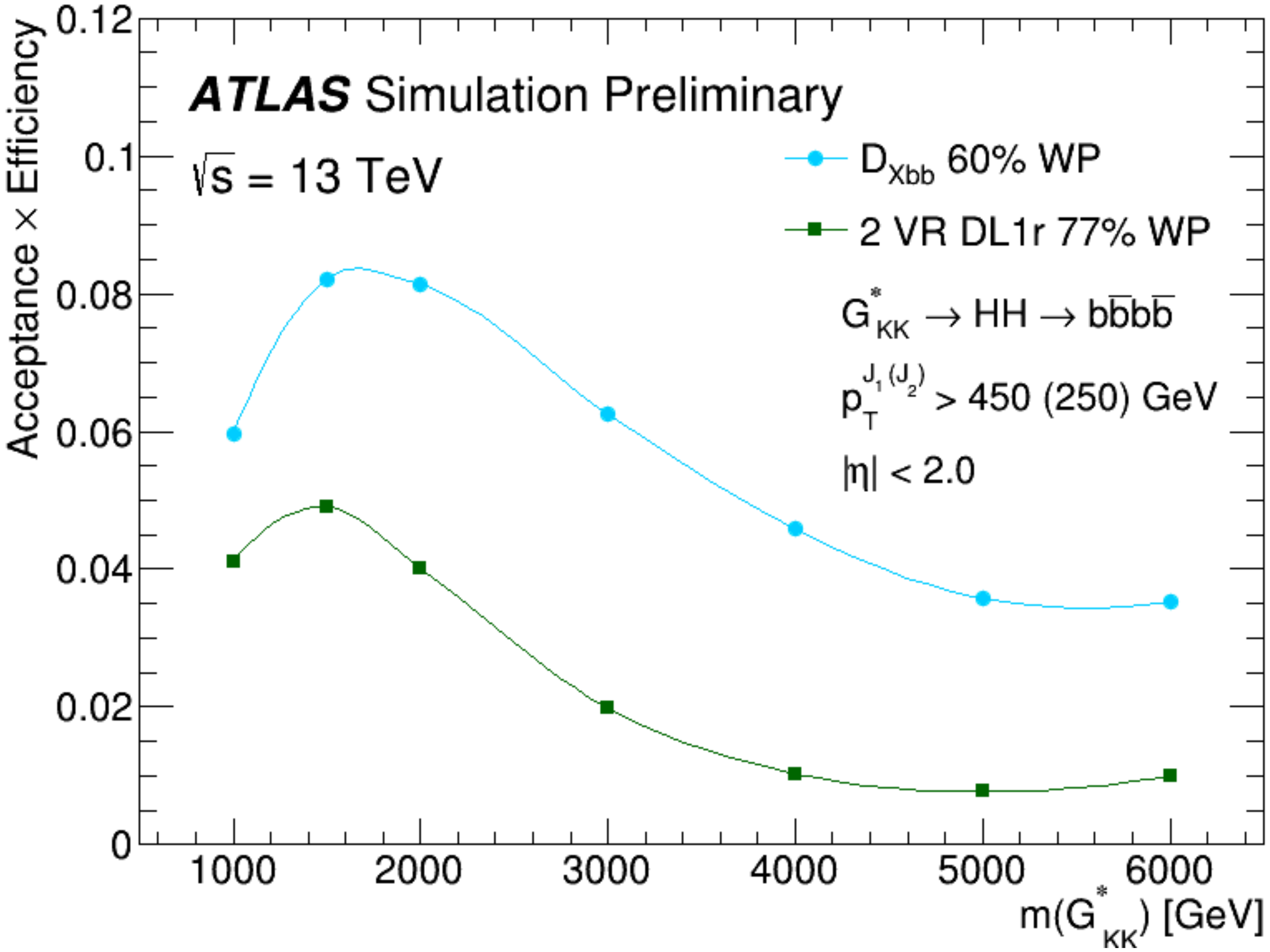
Xbb Tagger performance in G->H(bb)H(bb)

Cumulative acceptance x efficiency of signal, selecting two bb-tagged Large-R jets.

Acceptance is 50% to 350% better than two single-b tags in signal sample.

100% higher acceptance in background QCD multijet.





Xbb Tagger: performance

Table 1: Input variables used by the SVKine and JFKine, DL1 and DL1r algorithms.

Input	Variable	Description	SVKine	JFKine	DL1	DL1r
Kinematics	p_T	Jet p_T	✓	✓	✓	✓
	η	Jet $\|\eta\|$	✓	✓	✓	✓
IP2D, IP3D	$\log(P_b/P_{light})$	Likelihood ratio of the b-jet to light-flavour jet hypotheses			✓	✓
	$\log(P_b/P_c)$	Likelihood ratio of the b-jet to c-jet hypotheses			✓	✓
	$\log(P_c/P_{light})$	Likelihood ratio of the c-jet to light-flavour jet hypotheses			✓	✓
RNNIP	P_b	b-jet probability				✓
	P_c	c-jet probability				✓
	P_{light}	light-flavour jet probability				✓
SV1	$m(SV)$	Invariant mass of tracks at the secondary vertex assuming pion mass	✓		✓	✓
	$f_E(SV)$	Jet energy fraction of the tracks associated with the secondary vertex	✓		✓	✓
	$N_{TrkAtVtx}(SV)$	Number of tracks used in the secondary vertex	✓		✓	✓
	$N_{2TrkVtx}(SV)$	Number of two-track vertex candidates	✓		✓	✓
	$L_{xy}(SV)$	Transverse distance between the primary and secondary vertices	✓		✓	✓
	$L_{xyz}(SV)$	Distance between the primary and secondary vertices	✓		✓	✓
	$S_{xyz}(SV)$	Distance between the primary and secondary vertices divided by its uncertainty	✓		✓	✓
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})(SV)$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.	✓		✓	✓
JetFitter	$m(JF)$	Invariant mass of tracks from displaced vertices		✓	✓	✓
	$f_E(JF)$	Jet energy fraction of the tracks associated with the displaced vertices		✓	✓	✓
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})(JF)$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices		✓	✓	✓
	$S_{xyz}(JF)$	Significance of the average distance between PV and displaced vertices		✓	✓	✓
	$N_{TrkAtVtx}(JF)$	Number of tracks from multi-prong displaced vertices		✓	✓	✓
	$N_{2TrkVtx}(JF)$	Number of two-track vertex candidates (prior to decay chain fit)		✓	✓	✓
	$N_{1\text{-trk vertices}}(JF)$	Number of single-prong displaced vertices		✓	✓	✓
	$N_{\geq 2\text{-trk vertices}}(JF)$	Number of multi-prong displaced vertices		✓	✓	✓
	$L_{xyz}(2^{nd})(JF)$	Distance of 2^{nd} vertex from PV		✓	✓	✓
	$L_{xy}(2^{nd})(JF)$	Transverse displacement of the 2^{nd} vertex		✓	✓	✓
	$m_{Trk}(2^{nd})(JF)$	Invariant mass of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$E(2^{nd})(JF)$	Energy of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$f_E(2^{nd})(JF)$	Jet energy fraction of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$N_{TrkAtVtx}(2^{nd})(JF)$	Number of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$\eta_{trk}^{min,max,avg}(2^{nd})(JF)$	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex		✓	✓	✓

Table 1: Input variables used by the MV2 and the DL1 algorithms. The JetFitter c-tagging variables are only used by the DL1 algorithm.

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $\|\eta\|$
IP2D/IP3D	$\log(P_b/P_{light})$	Likelihood ratio between the b-jet and light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{light})$	Likelihood ratio between the c-jet and light-flavour jet hypotheses
SV1	$m(SV)$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(SV)$	Energy fraction of the tracks associated with the secondary vertex
	$N_{TrkAtVtx}(SV)$	Number of tracks used in the secondary vertex
	$N_{2TrkVtx}(SV)$	Number of two-track vertex candidates
	$L_{xy}(SV)$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(SV)$	Distance between the primary and the secondary vertex
	$S_{xyz}(SV)$	Distance between the primary and the secondary vertex divided by its uncertainty
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})(SV)$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JetFitter	$m(JF)$	Invariant mass of tracks from displaced vertices
	$f_E(JF)$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})(JF)$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(JF)$	Significance of the average distance between PV and displaced vertices
	$N_{TrkAtVtx}(JF)$	Number of tracks from multi-prong displaced vertices
	$N_{2TrkVtx}(JF)$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{1\text{-trk vertices}}(JF)$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-trk vertices}}(JF)$	Number of multi-prong displaced vertices
JetFitter c-tagging	$L_{xyz}(2^{nd}/3^{rd}vtx)(JF)$	Distance of 2^{nd} or 3^{rd} vertex from PV
	$L_{xy}(2^{nd}/3^{rd}vtx)(JF)$	Transverse displacement of the 2^{nd} or 3^{rd} vertex
	$m_{Trk}(2^{nd}/3^{rd}vtx)(JF)$	Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex
	$E_{Trk}(2^{nd}/3^{rd}vtx)(JF)$	Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex
	$f_E(2^{nd}/3^{rd}vtx)(JF)$	Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex
	$N_{TrkAtVtx}(2^{nd}/3^{rd}vtx)(JF)$	Number of tracks associated with 2^{nd} or 3^{rd} vertex
	$Y_{trk}^{min}, Y_{trk}^{max}, Y_{trk}^{avg}(2^{nd}/3^{rd}vtx)(JF)$	Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex