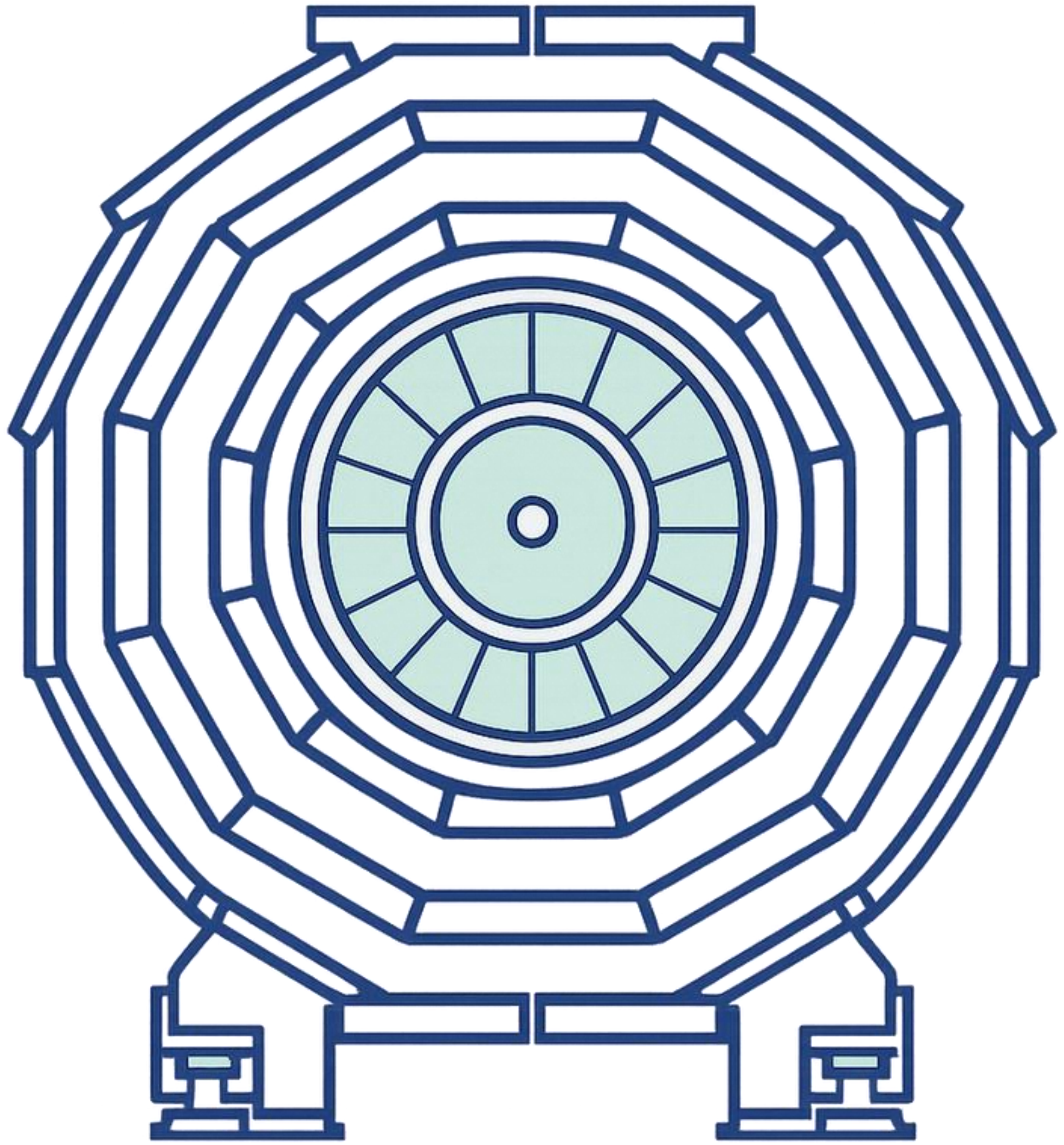


B-TAGGING FOR RUN 3 NMSSM ANALYSIS AT CMS

EMMA CLARASÓ BATLLORI

NExT PhD Workshop

26/08/2026

Science and Technology Facilities Council

OUTLINE

1 **INTRODUCTION**
NMSSM Analysis
b-tagging

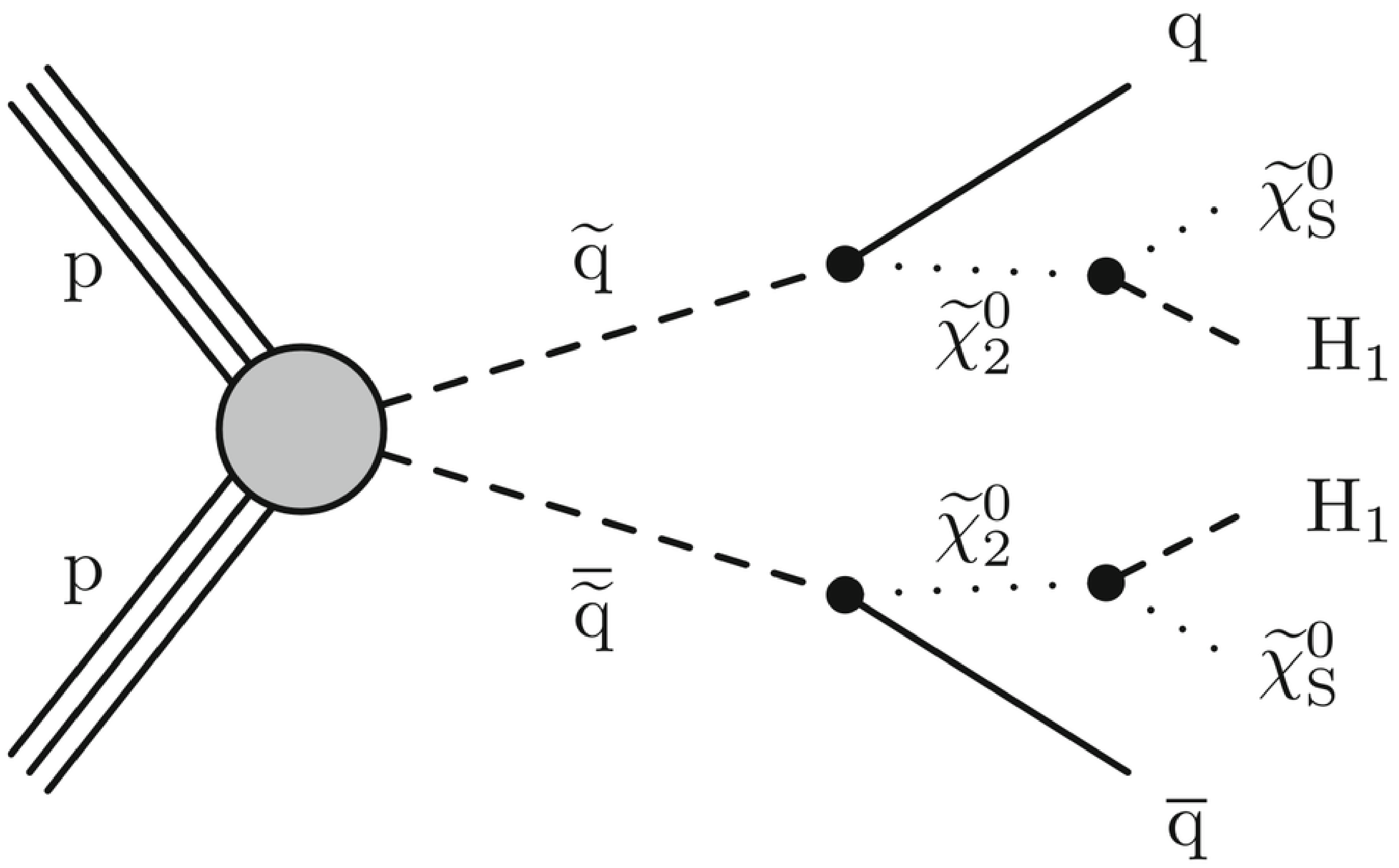
2 **B-TAGGING PERFORMANCE**
Mass Dependence
GloParT
ΔR_{bb} Dependence

3 **B-TAGGING SCALE FACTORS**
μ-tagged Method
Mutag-calib Framework

4 **CONCLUSIONS & FURTHER WORK**

INTRODUCTION
NMSSM

- NMSSM predicts an extended Higgs Sector with 7 Higgs particles
- The focus of this analysis is the lightest CP-even Higgs boson (H_1)
- The signature is a pair of H_1 bosons produced in SUSY cascade decays, which arise from the pair production of squarks and gluinos



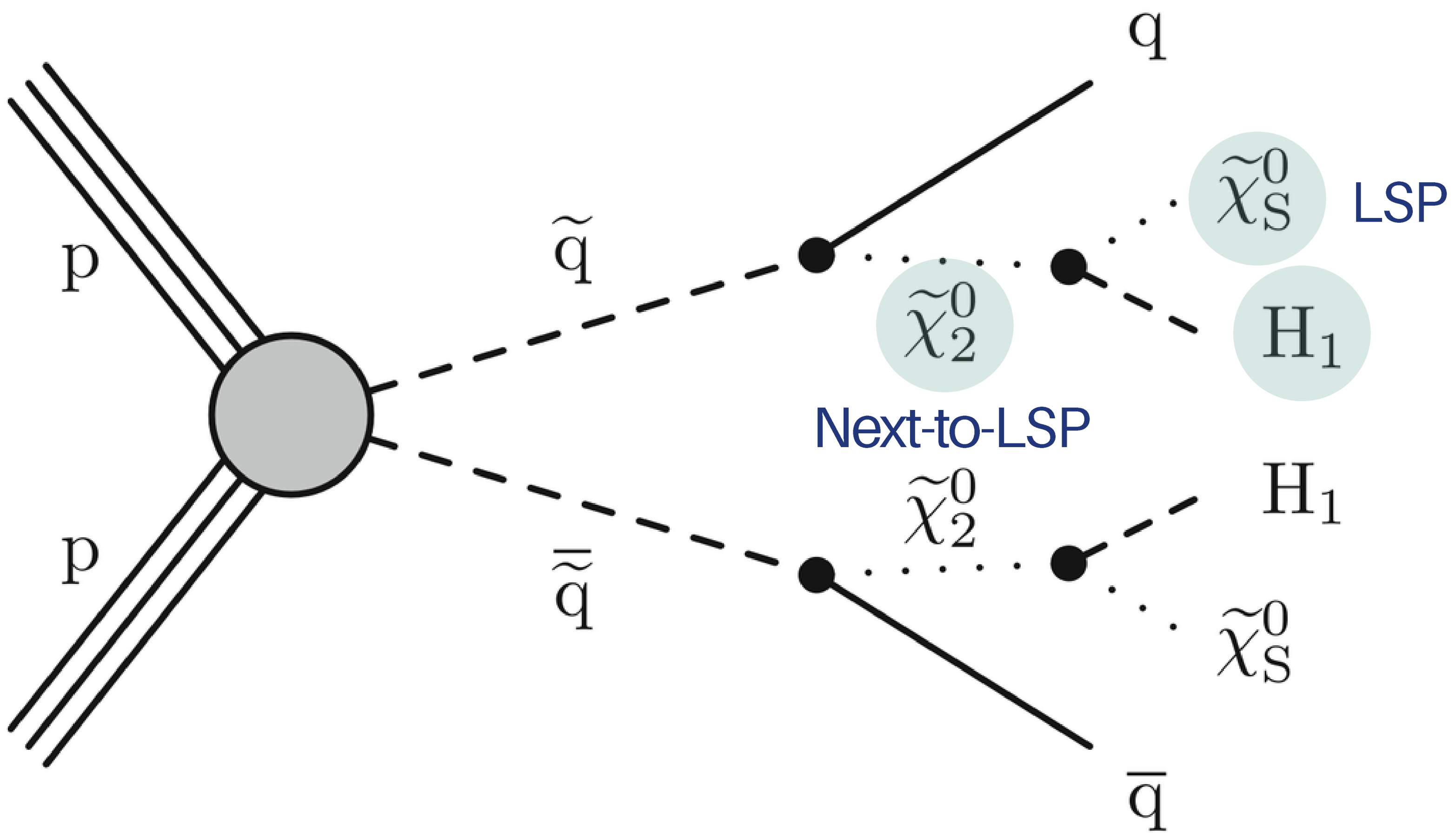
INTRODUCTION
NMSSM

$\tilde{\chi}_S^0$ mass $\ll H_1$ mass

Phase space for $\tilde{\chi}_2^0 \rightarrow \tilde{\chi}_S^0 + H_1$ small

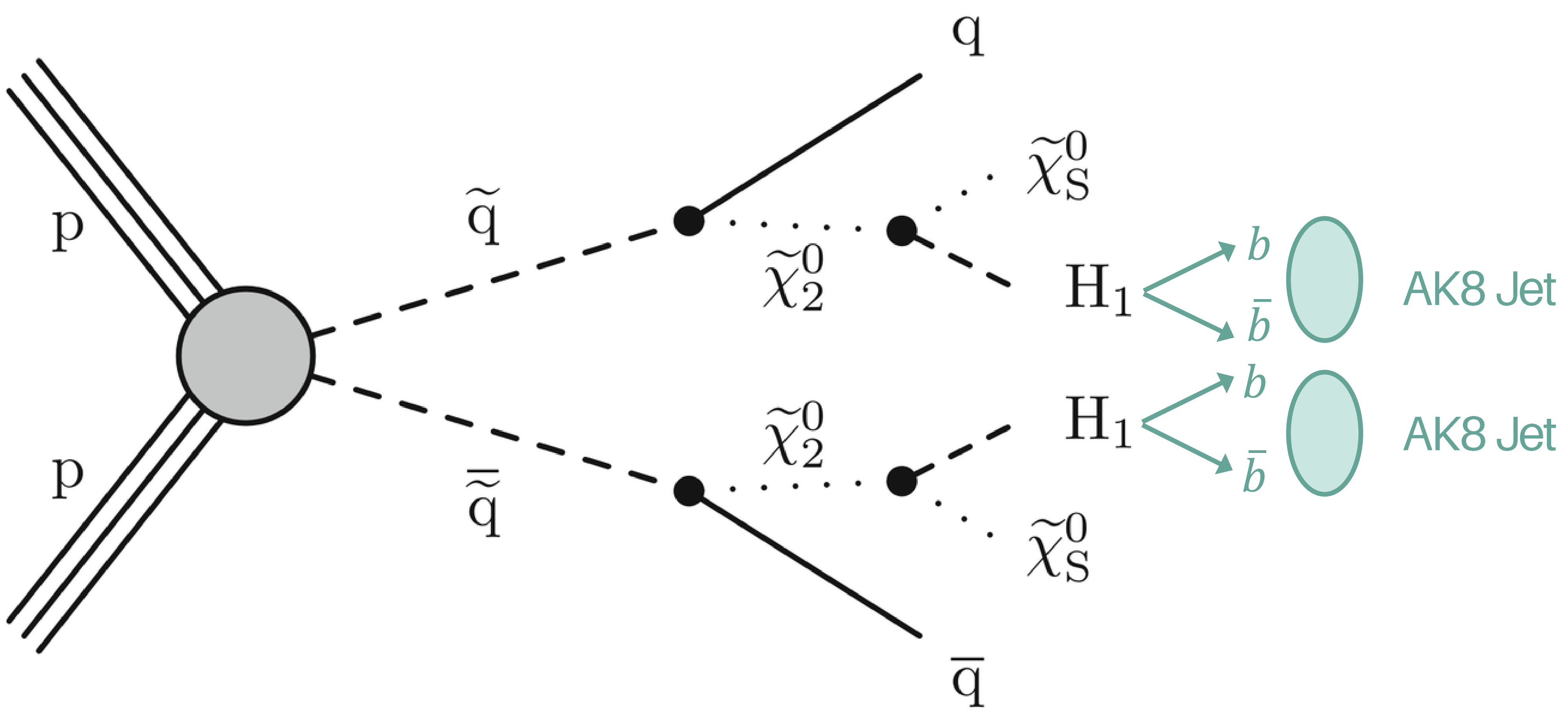
→ H_1 carries much larger momentum than the $\tilde{\chi}_S^0$

→ Small p_T^{miss} scenario



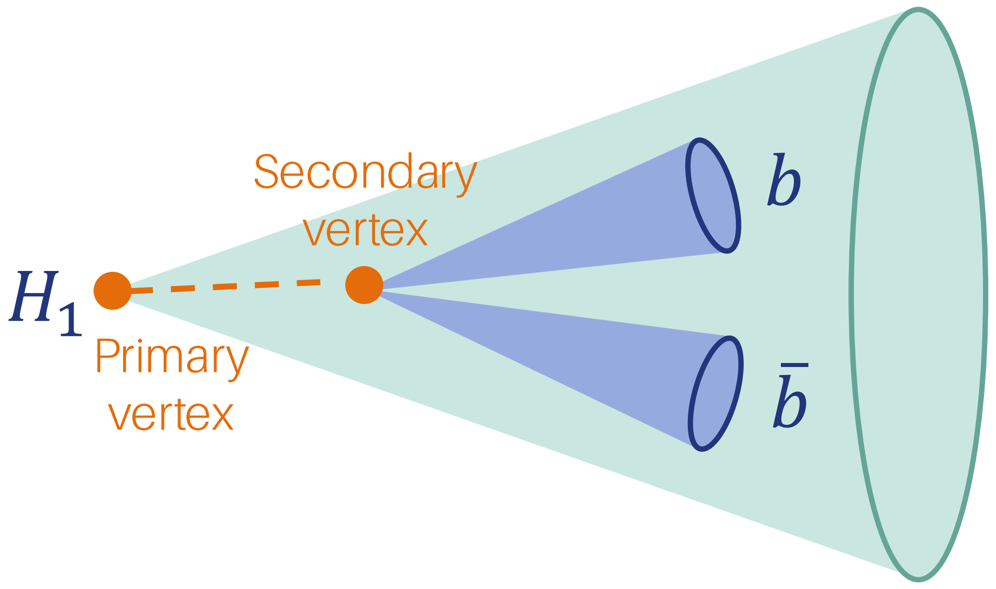
INTRODUCTION
NMSSM

- Search for 2 highly-boosted H_1 bosons that decay into $b\bar{b}$ pairs, reconstructed as 2 large radius (AK8) jets using substructure techniques → **(double) b-tagging is important!**

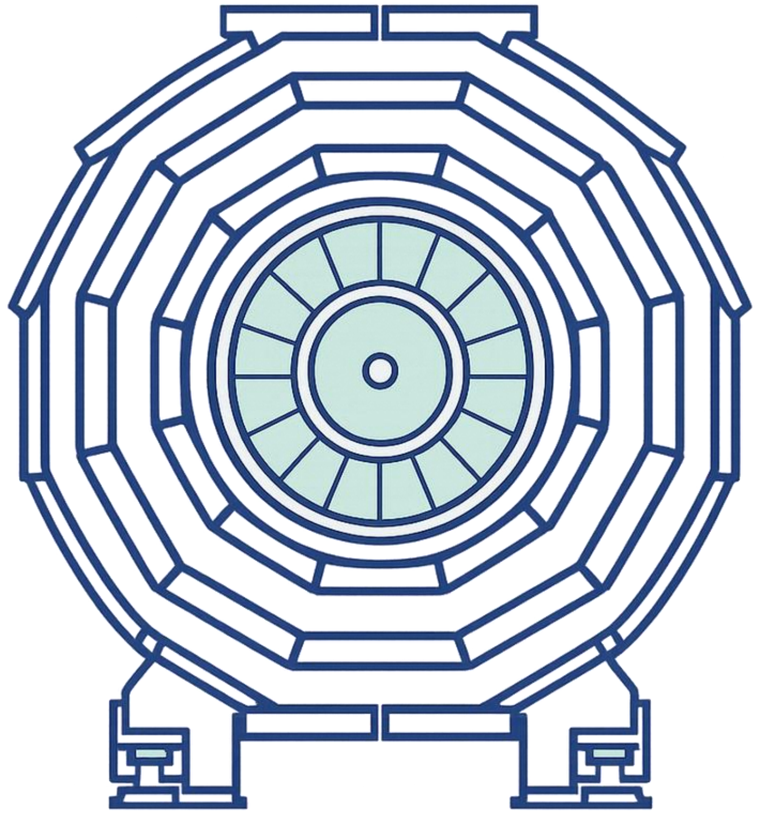


INTRODUCTION
B-TAGGING

- Identifying jets originating from b-quarks using characteristic B-hadron signatures:
  - Long lifetime → displaced tracks and secondary vertices
  - Relatively large B-hadron mass
  - Semileptonic decays → possible soft leptons inside the jet
- Jet taggers are ML-based these days

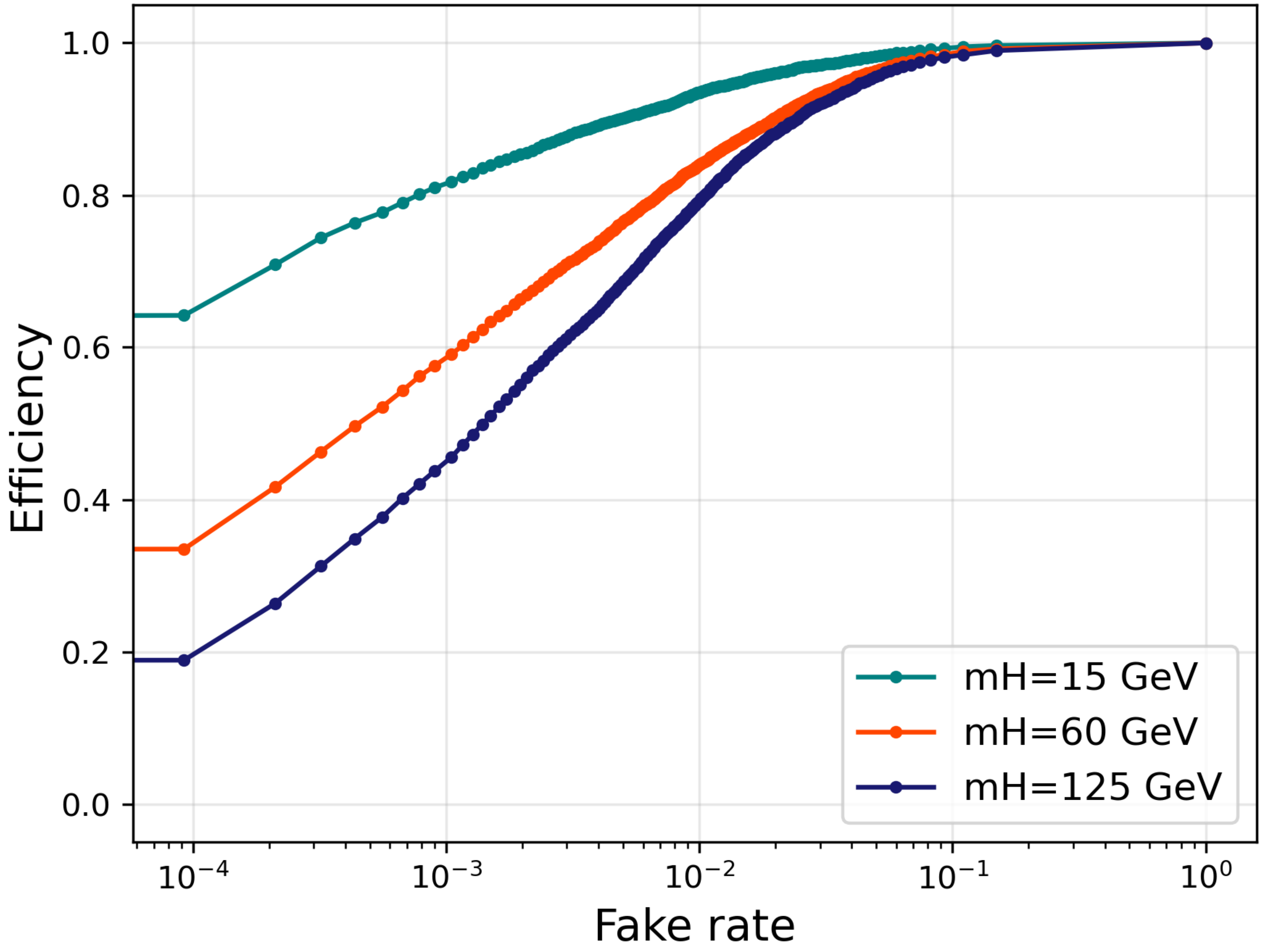


- For a boosted $H_1 \to b\bar{b}$ decay, both b-quarks can be reconstructed inside a single large-radius AK8 jet.
- This is not conventional single b-tagging: the tagger must identify a two-prong $b\bar{b}$ structure within one jet and distinguish it from QCD jets.



B-TAGGING PERFORMANCE ACROSS MASSES

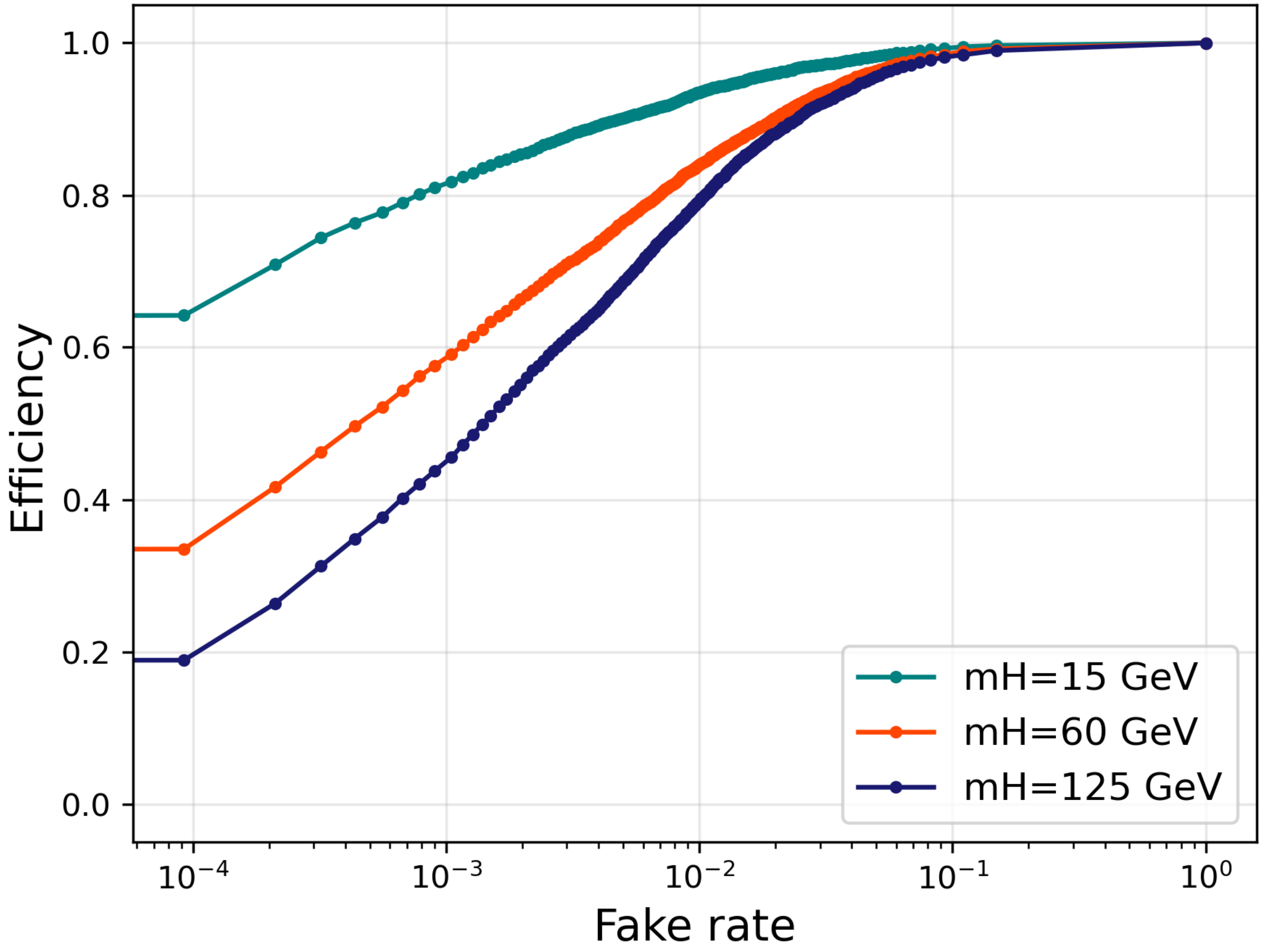
B-TAGGING PERFORMANCE
MASS DEPENDENCE



- Analysis performs search of H_1 bosons in mass range 10 -125 GeV
- Study double b-tagging performance across masses

B-TAGGING PERFORMANCE

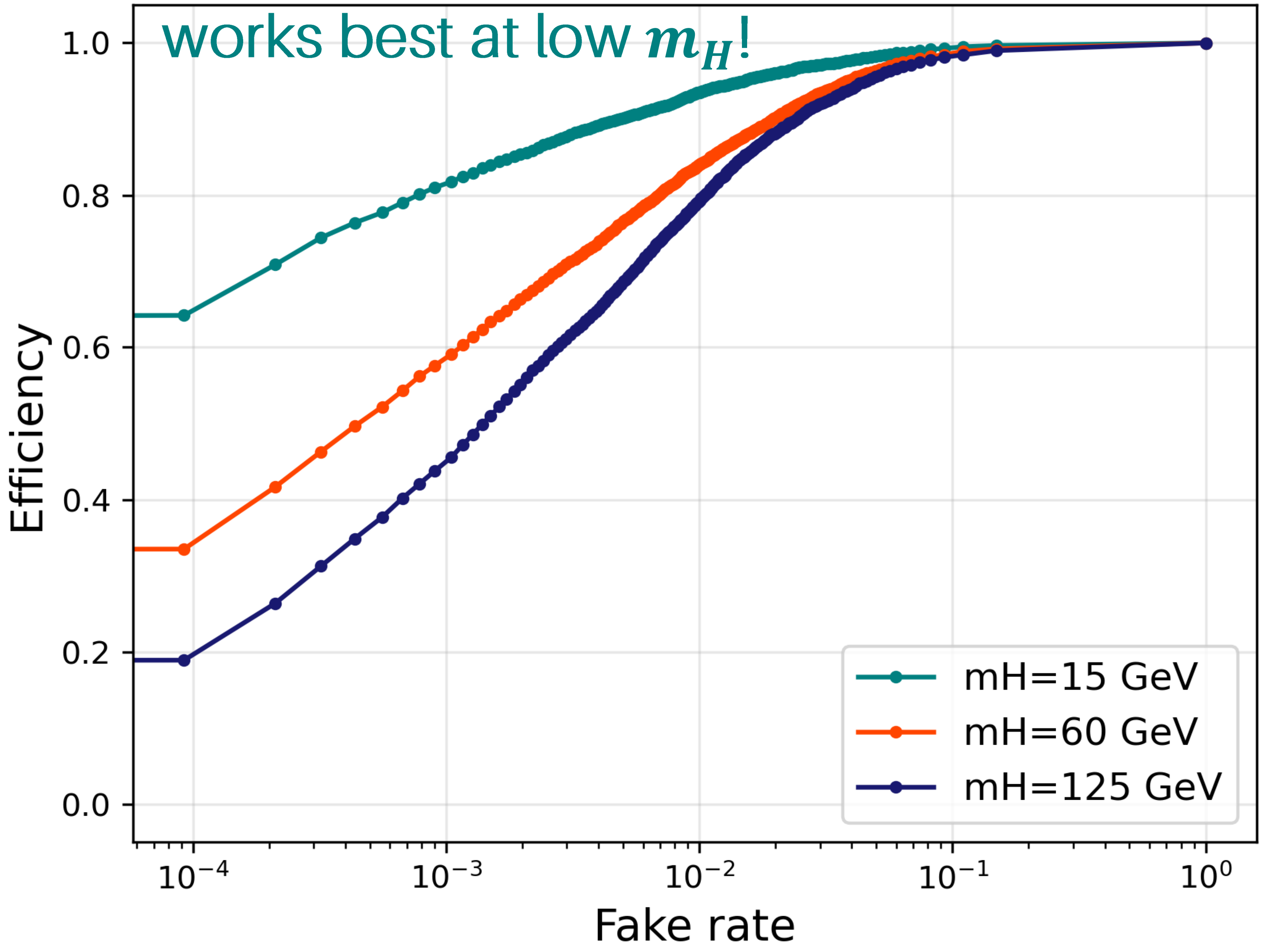
MASS DEPENDENCE



- Output of the tagger is a score between 0 and 1.
- For a given score threshold:
  - **Efficiency:** Fraction of signal jets above threshold.
  - **Fake rate:** Fraction of QCD background jets above same threshold.
- Scan over thresholds to produce ROC curve

B-TAGGING PERFORMANCE
MASS DEPENDENCE

Found that tagger works best at low m_H!



- Output of the tagger is a score between 0 and 1.
- For a given score threshold:
  - **Efficiency:** Fraction of signal jets above threshold.
  - **Fake rate:** Fraction of QCD background jets above same threshold.
- Scan over thresholds to produce ROC curve

B-TAGGING PERFORMANCE

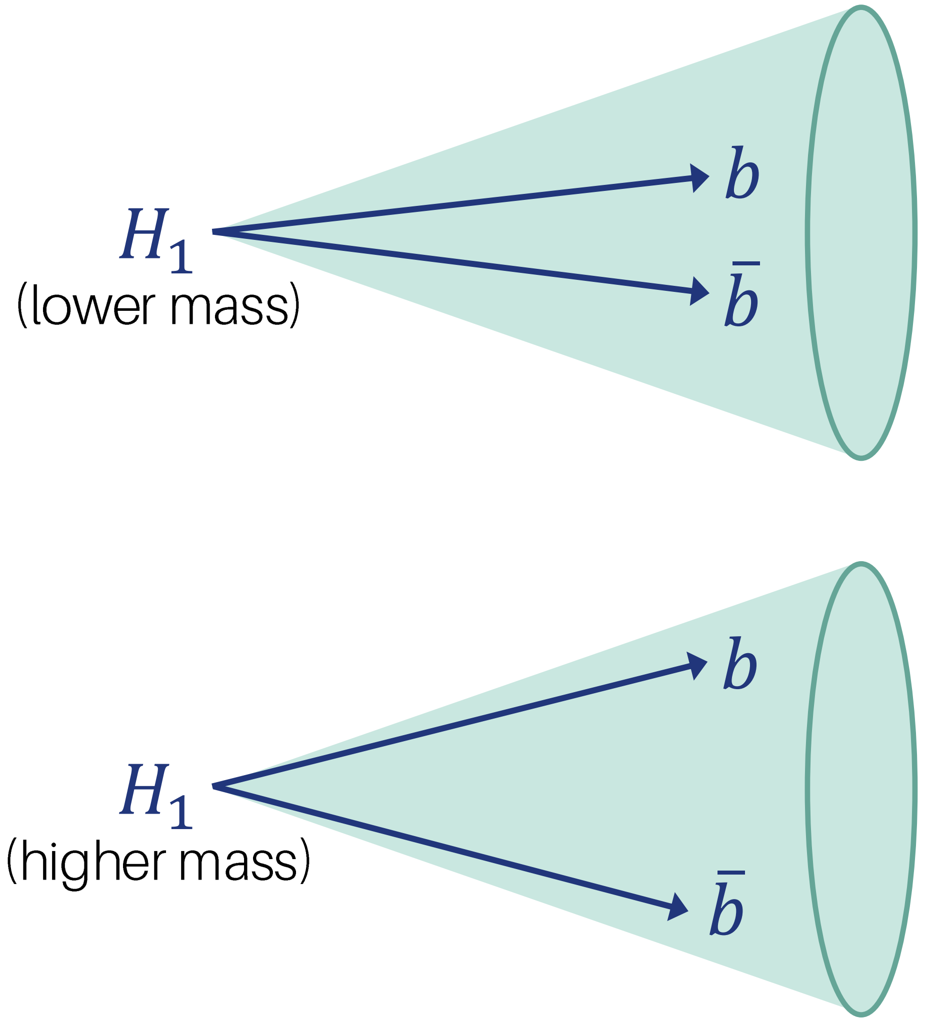
GLOPART

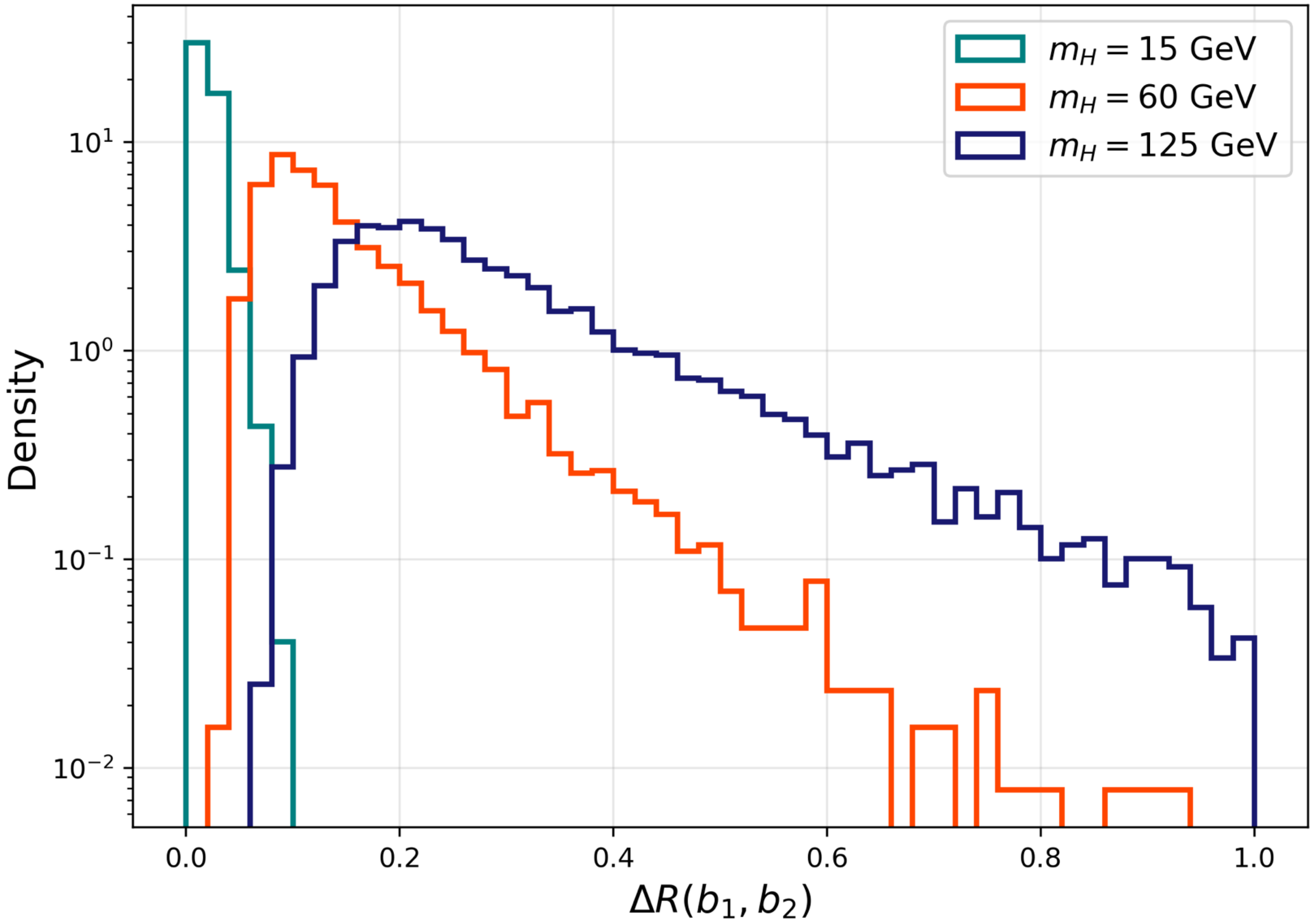
- Global Particle Transformer (GloParT) is a transformer-based jet tagger that uses jet constituents to identify boosted $X \rightarrow b\bar{b}$ decays.
- The tagger is trained to be independent of jet mass
  - Mass not used as input
  - Training reweighted in m_{SD} and p_{T} to remove their discriminating power
- Despite this, substructure observables can still be indirectly affected due to their correlation with jet mass and p_{T}

The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3, https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

ΔR_{bb} DEPENDENCE

- In particular, the angular separation between the b-quarks (ΔR_{bb}) differs significantly across masses

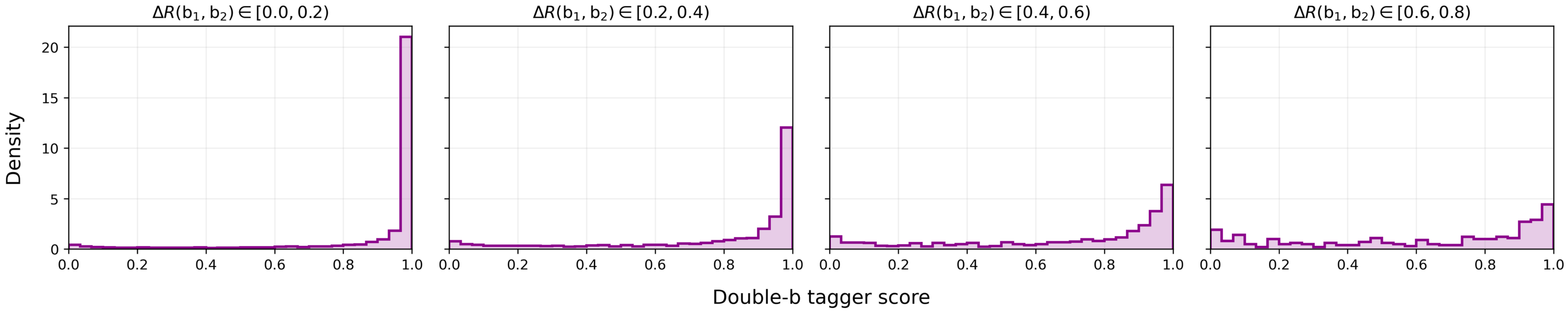




B-TAGGING PERFORMANCE

ΔR_{bb} DEPENDENCE

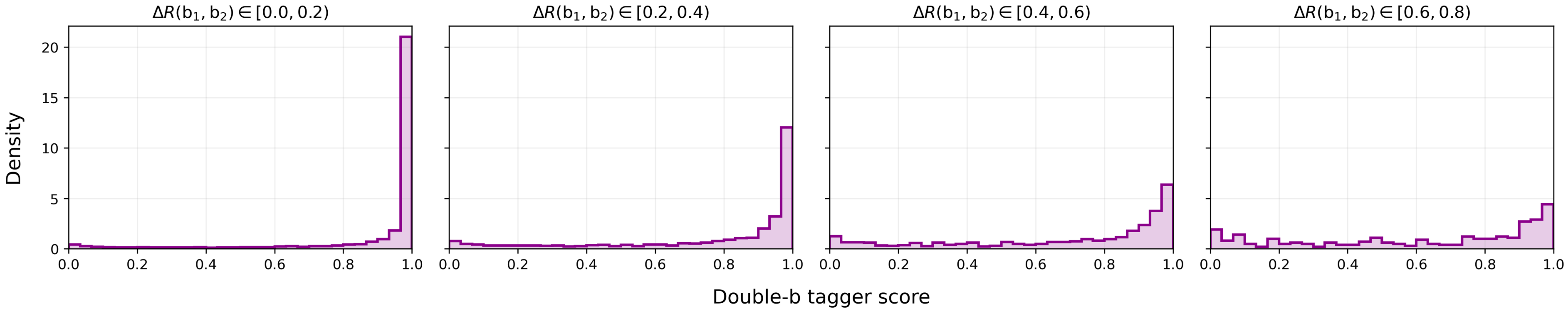
- Look at tagger performance as a function of ΔR_{bb}



B-TAGGING PERFORMANCE

ΔR_{bb} DEPENDENCE

- Look at tagger performance as a function of ΔR_{bb}

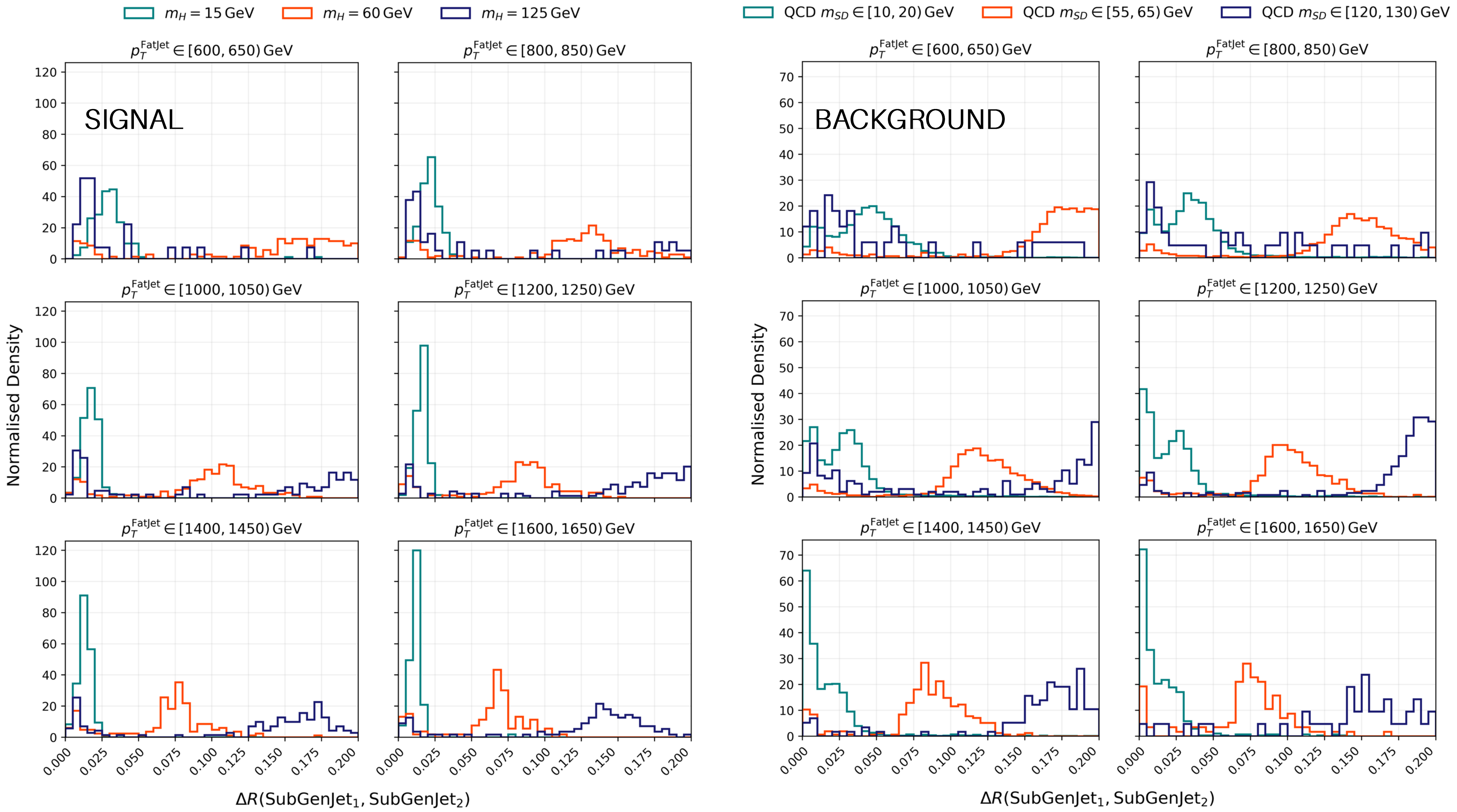


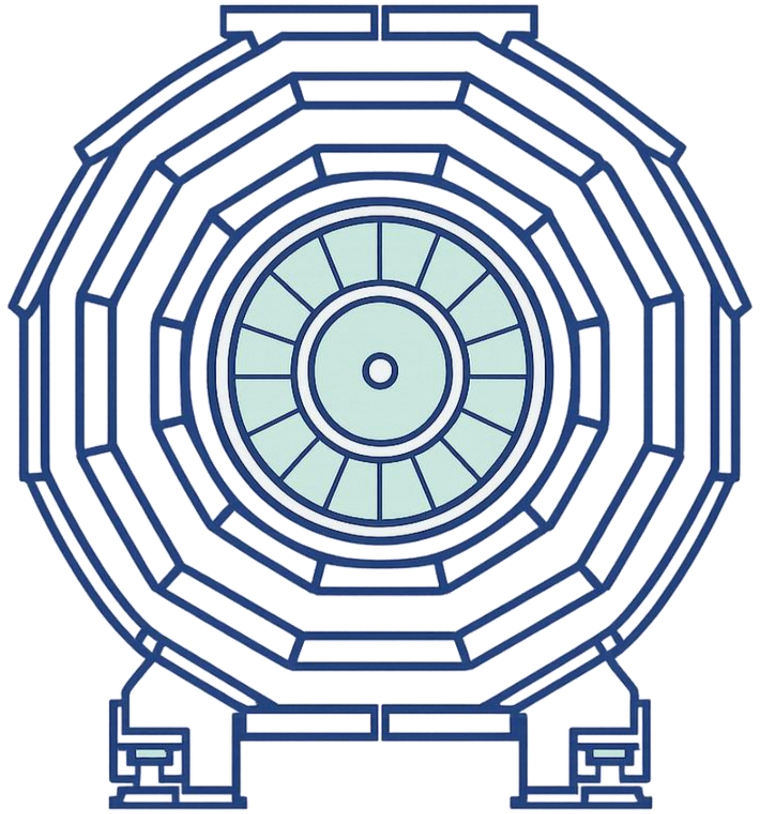
- Look at distributions for signal and QCD

ΔR_{bb} DEPENDENCE

- Cannot define ΔR_{bb} for background sample → Makes comparison difficult
- Find a proxy variable:
  - Can be used to select roughly the same sample of jets in each bin as ΔR_{bb}
  - Can be equally defined for signal and QCD
- Try angular separation between SubGenJets ($\boldsymbol{\Delta R_{SJ}}$)
  - Reasonable in the lower ΔR bin

B-TAGGING PERFORMANCE
ΔR_{bb} DEPENDENCE





B-TAGGING SCALE FACTOR DERIVATION

B-TAGGING SCALE FACTORS
μ-TAGGED METHOD

- GloParT trained only on simulation → Need to derive scale factors to correct for differences in the b-tagging response between data and simulation.
  $$\mathrm{SF} = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}}$$
- The method calibrates the $X \to b\bar{b}$ signal jets using proxy jets from gluon splitting $b\bar{b}$ jets that contain a soft muon within their respective cones.
- Hadronised final state initiated from the decay of a b-quark has a 20% probability of including a muon → requirement provides a good handle to select a pure sample of heavy-flavour jets.

The CMS Collaboration, Performance of heavy-flavour jet identification in Lorentz-boosted topologies in proton-proton collisions at √ s = 13 TeV, https://arxiv.org/pdf/2510.10228

μ-TAGGED METHOD

	Selection	Tagger WPs	Flavour Composition	Efficiency calculation
Monte Carlo	Common selection applied at trigger and offline levels: • Event selection via jet + muon trigger • Jet selection with kinematic and muon-in-jet requirements	Define pass / fail regions by applying tagger WPs	Jets matched to truth-level hadrons to identify their flavour (b, c, light).	$\epsilon = \frac{N_{pass,b}}{N_{pass,b} + N_{fail,b}}$
Data			Simulation-derived Σm_{SV} templates are fitted to the data to extract the flavour composition. $\text{Data} = N_b T_b + N_c T_c + N_l T_l$	

μ-TAGGED METHOD

	Selection	Tagger WPs	Flavour Composition	Efficiency calculation
Monte Carlo	Common selection applied at trigger and offline levels: • Event selection via jet + muon trigger • Jet selection with kinematic and muon-in-jet requirements	Define pass / fail regions by applying tagger WPs	Jets matched to truth-level hadrons to identify their flavour (b, c, light).	$\epsilon = \frac{N_{pass,b}}{N_{pass,b} + N_{fail,b}}$
Data			Simulation-derived Σm_{SV} templates are fitted to the data to extract the flavour composition. $\text{Data} = N_b T_b + N_c T_c + N_l T_l$	

B-TAGGING SCALE FACTORS
μ-TAGGED METHOD

	Selection	Tagger WPs	Flavour Composition	Efficiency calculation
Monte Carlo	Common selection applied at trigger and offline levels: • Event selection via jet + muon trigger • Jet selection with kinematic and muon-in-jet requirements	Define pass / fail regions by applying tagger WPs	Jets matched to truth-level hadrons to identify their flavour (b, c, light).	$\epsilon = \frac{N_{pass,b}}{N_{pass,b} + N_{fail,b}}$
Data			Simulation-derived Σm_{SV} templates are fitted to the data to extract the flavour composition. $\mathrm{Data} = N_b T_b + N_c T_c + N_l T_l$	

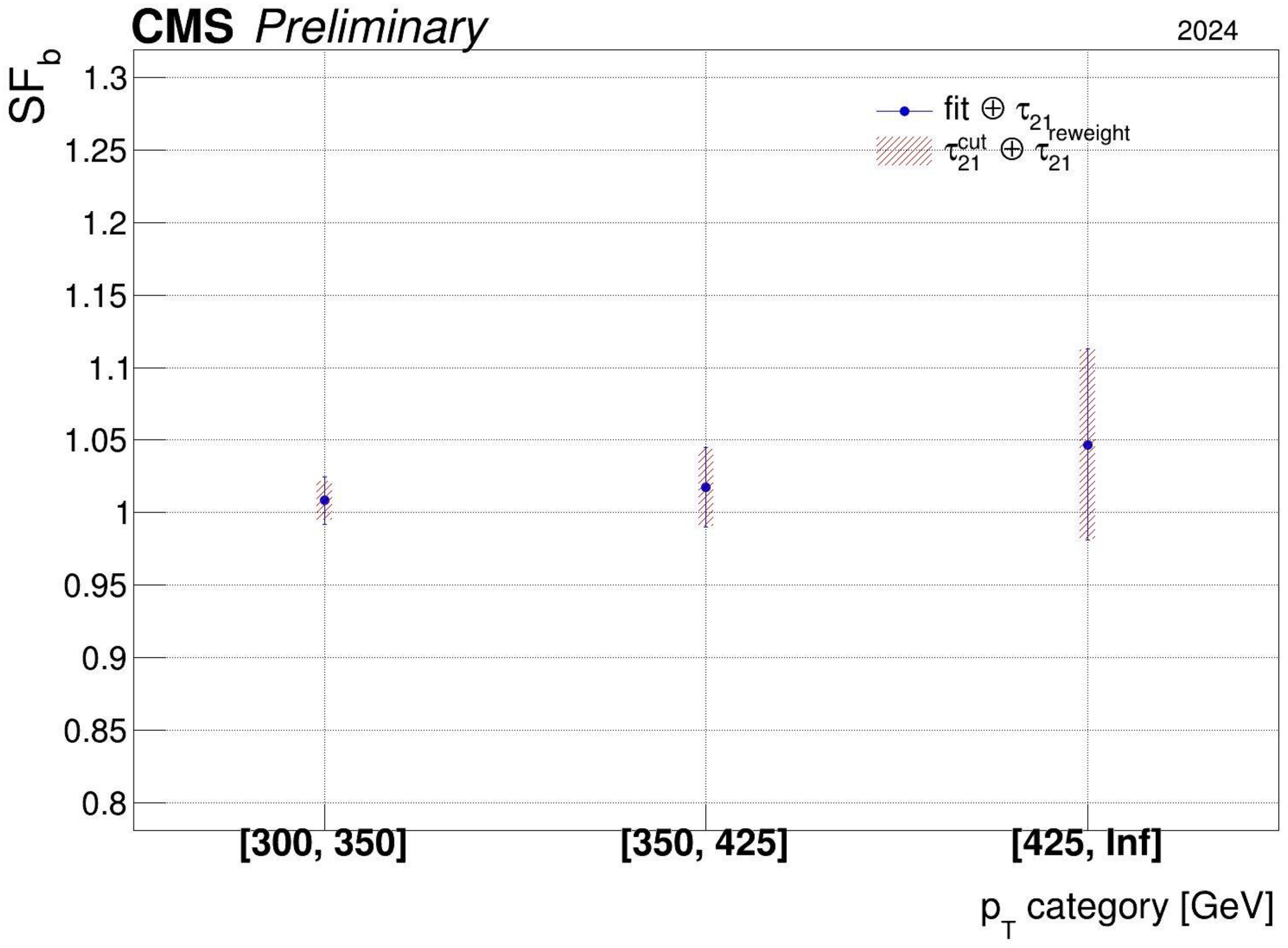
B-TAGGING SCALE FACTORS

μ-TAGGED METHOD

	Selection	Tagger WPs	Flavour Composition	Efficiency calculation
Monte Carlo	Common selection applied at trigger and offline levels: • Event selection via jet + muon trigger • Jet selection with kinematic and muon-in-jet requirements	Define pass / fail regions by applying tagger WPs	Jets matched to truth-level hadrons to identify their flavour (b, c, light).	$\epsilon = \frac{\mathrm{N_{pass,b}}}{\mathrm{N_{pass,b}} + \mathrm{N_{fail,b}}}$
Data			Simulation-derived Σm_{SV} templates are fitted to the data to extract the flavour composition. $\mathrm{Data} = \mathrm{N_b T_b} + \mathrm{N_c T_c} + \mathrm{N_l T_l}$	

MUTAG-CALIB FRAMEWORK

year	category p_T [GeV]	$SF_{nominal}$	err_{fit}	τ_{21}^{cut}	$\tau_{21}^{reweight}$	σ_{tot}
2024	[300, 350]	1.008	0.009	0.013	0.002	0.017
2024	[350, 425]	1.017	0.007	0.019	0.018	0.027
2024	[425, ∞]	1.047	0.009	0.030	0.058	0.066



- Method implemented in for di-Higgs analyses for now
- Initial validation step: run the full framework for $HHbb\gamma\gamma$
  - Asked the author for obtained results
  - Now have perfect agreement!
- Start working on adapting the framework to our analysis

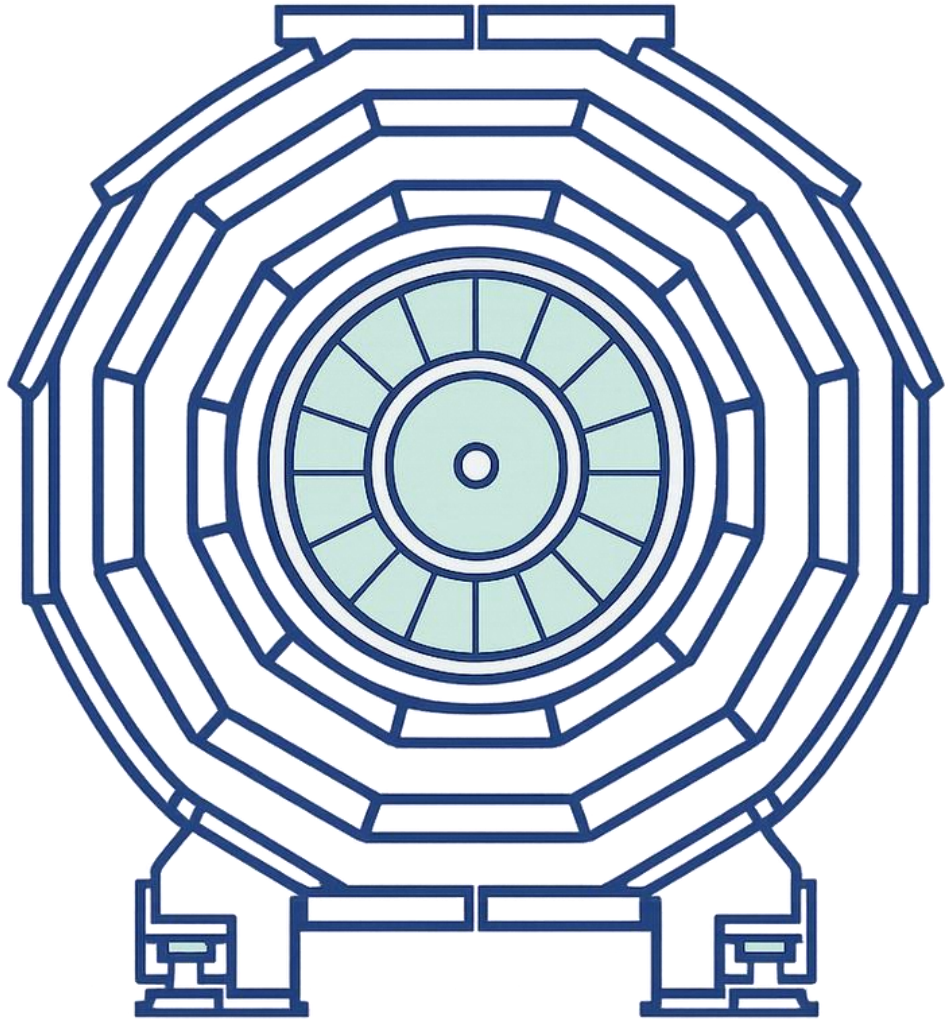
CONCLUSIONS & FURTHER WORK

- Analysis presents a search of light Higgs bosons with small missing transverse momentum, which deviates from traditional SUSY searches
- Identification of H_1 jets relies heavily on double b-tagging, which is a crucial part of the analysis
- Focus on two aspects:
  - b-tagging performance study

Uncovered mass dependence of the tagger, related to ΔR_{bb}	→	Currently looking into accessing relevant MC information for QCD sample

  - Scale factor derivation

Validated the framework for $HHbb\gamma\gamma$ analysis	→	Adapt framework to NMSSM analysis and test its robustness



THANK YOU!

BACKUP

GLOPART SPECIFICATIONS

Table 3: Summary of the GloParT training setup and comparison among the three model versions: model architecture.

	ParT for $H \to WW$ (GloParT v1)	GloParT v2	GloParT v3
Model architecture			
Input-feature embedding dimensions	(128, 512, 128)	(256, 1024, 256)	(256, 1024, 256)
Pairwise-feature embedding dimensions	(64, 64, 64)	(64, 64, 64)	(64, 64, 64)
Transformer embedding dimension	128	256	256
Transformer blocks (particle attention block + class attention block [7])	$8+2$	$8+2$	$10+2$
Attention heads	8	16	16
Model parameters	2.3 M	10.2 M	16.3 M
FLOPs	0.43 GMac	1.73 GMac	1.60 GMac
Training configuration			
Batch size	1024	256	640
Samples processed per epoch	7.68 M	7.68 M	7.68 M
Training epochs	50	50	100
Activation function	GELU	GELU	GELU/SwiGLU
Training and validation dataset size	52.4 M	108 M	303 M
Optimizer	Ranger	Ranger	Ranger
Learning rate	2×10^{-3}	2×10^{-3}	1.2×10^{-3}

The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3 (slide 34)
https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

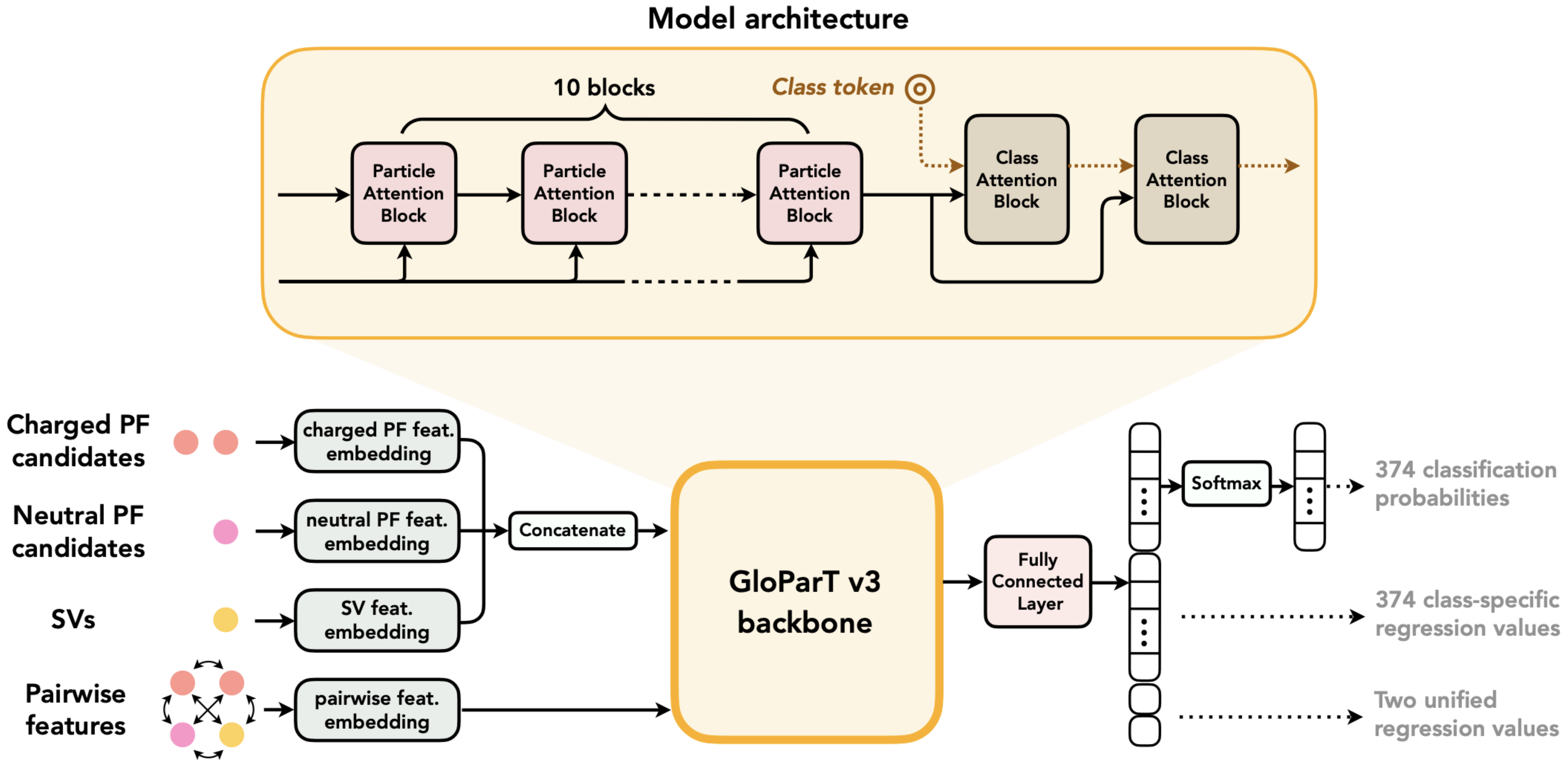
GLOPART SPECIFICATIONS

Table 3: Summary of the GloParT training setup and comparison among the three model versions, continued: training configuration and input representation.

	ParT for $H \to WW$ (GloParT v1)	GloParT v2	GloParT v3
Input representation			
Input groups	Two groups: charged and neutral PF candidates, and SVs	Three groups: charged PF candidates, neutral PF candidates, and SVs. Lost tracks are included in the charged-PF-candidate group.	
Maximum token multiplicity	$(128, 10)$	$(90, 60, 10)$	
Input features	25 PF-candidate features and 11 SV features, as described in Ref. [11]	Charged and neutral PF candidates are represented separately. Track-related features are used only for charged PF candidates, and hit-pattern variables are added to the charged-candidate feature set. The charged-candidate, neutral-candidate, and SV feature sets contain 30, 7, and 11 features, respectively.	
Pairwise features	$\ln \Delta$, $\ln k_T$, $\ln z$, and $\ln m^2$ [7]	Same as GloParT v1	In addition to the four variables, two normalized features are also included: $k_{T,ij}/\sum_\ell p_{T,\ell}$ and $m_{ij}/m(\sum_\ell p_\ell)$

The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3 (slide 35)
https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

GLOPART SPECIFICATIONS



The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3 (slide 35)
https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

GLOPART SPECIFICATIONS

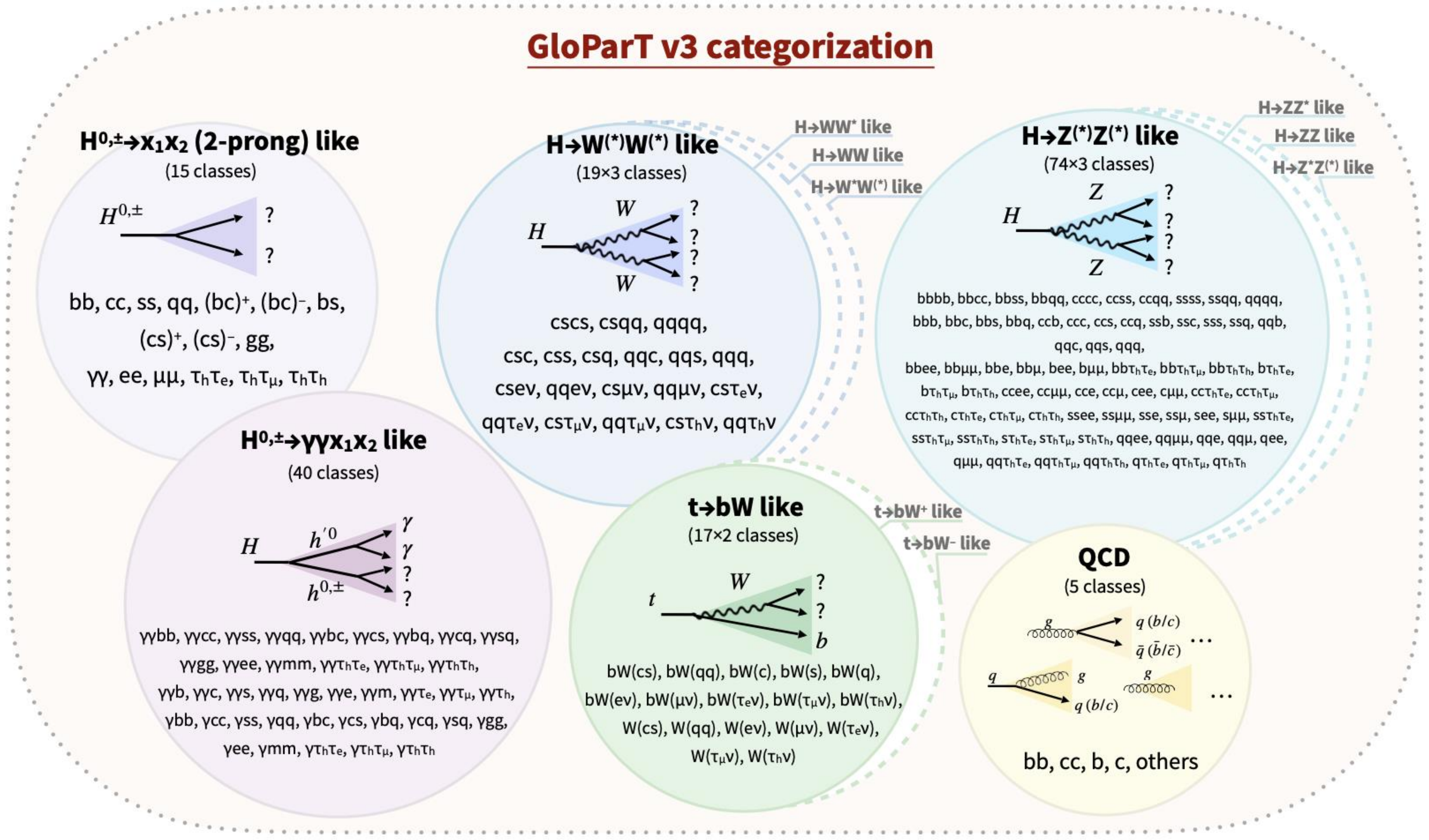


Figure 24: GloParT v3 is trained with 374 jet classes, including conventional two-prong signatures; multiprong signatures categorized by final state and the quark or lepton flavour originating from $H \to WW$-like and $H \to ZZ$-like processes, as well as processes containing two photons and two partons or leptons; top-like three-prong jet signatures; and QCD jet categories. Relative to GloParT v2, the categorization of two-prong signatures is further refined, and additional classes containing photons are introduced.

The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3 (slide 39)
https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

GLOPART DISCRIMINANT

Discriminant definition

For each tagging scenario, a discriminant is constructed by selecting one or more signal classes, denoted by $\mathcal{S}$, from the 374 GloParT v3 output categories. The discriminant is defined as

$$D_{\mathcal{S}\ \mathrm{vs\ QCD}} = \frac{\sum_{i\in\mathcal{S}} \alpha_i\, p_i}{\sum_{i\in\mathcal{S}} \alpha_i\, p_i + \sum_{i\in\mathrm{QCD}} p_i},$$

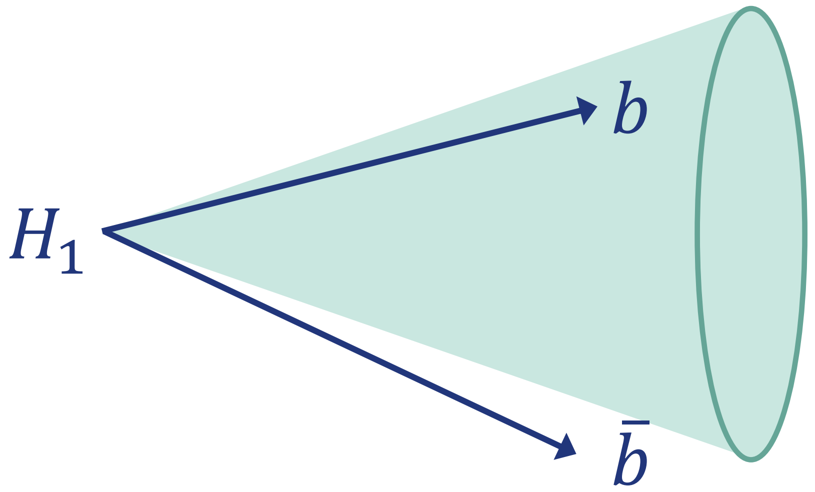
where $\mathcal{S}$ denotes the set of signal classes, p_i is the predicted probability for class i, and α_i is an optional weighting factor applied to each signal class. As discussed in Ref. [21], this definition naturally exploits the multiclass and fine-grained classification structure, without introducing an intrinsic loss of discrimination power.

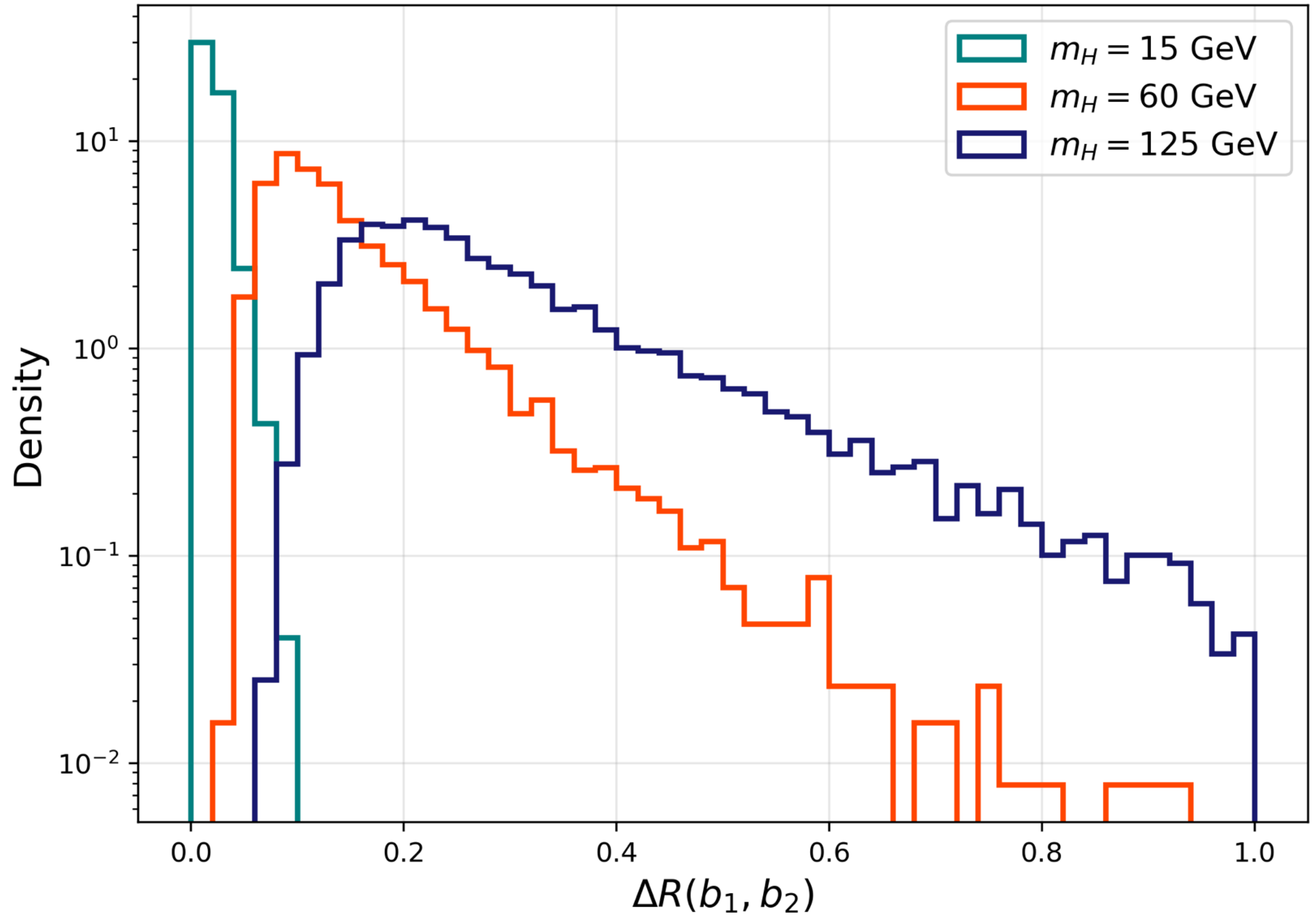
For $\mathrm{H}\to\mathrm{b\bar{b}}$ and $\mathrm{H}\to\mathrm{c\bar{c}}$ tagging, the signal classes are chosen as the corresponding $\mathrm{b\bar{b}}$ and $\mathrm{c\bar{c}}$ categories. For $\mathrm{H}\to\mathrm{q\bar{q}}$ ($\mathrm{q}=\mathrm{u,d,s}$) tagging, the signal score in the numerator is defined as $p_{\mathrm{q\bar{q}}} \equiv \frac{1}{3}p_{\mathrm{s\bar{s}}} + \frac{2}{3}p_{\mathrm{u\bar{u}/d\bar{d}}}$, where $\mathrm{s\bar{s}}$ and $\mathrm{u\bar{u}/d\bar{d}}$ are the two signal classes. For hadronic W tagging, the signal score is defined as $p_{\mathrm{c\bar{s}}} + p_{\mathrm{\bar{c}s}} + p_{\mathrm{q\bar{q}}}$, where the combination is chosen according to the available GloParT output scores. For fully-hadronic $\mathrm{H}\to\mathrm{WW}$ tagging, the signal definition includes the sum of the output probability nodes corresponding to the $\mathrm{X}\to\mathrm{WW}\to 4\mathrm{q}$ decay modes and the associated three-quark topologies, in which one quark falls outside the reconstructed jet [16].

The CMS Collaboration, The Global Particle Transformer for boosted-jet tagging and mass regression: Development and performance from v1 to v3 (slide 8)
https://cds.cern.ch/record/2966239/files/DP2026_104.pdf

b-quark containment

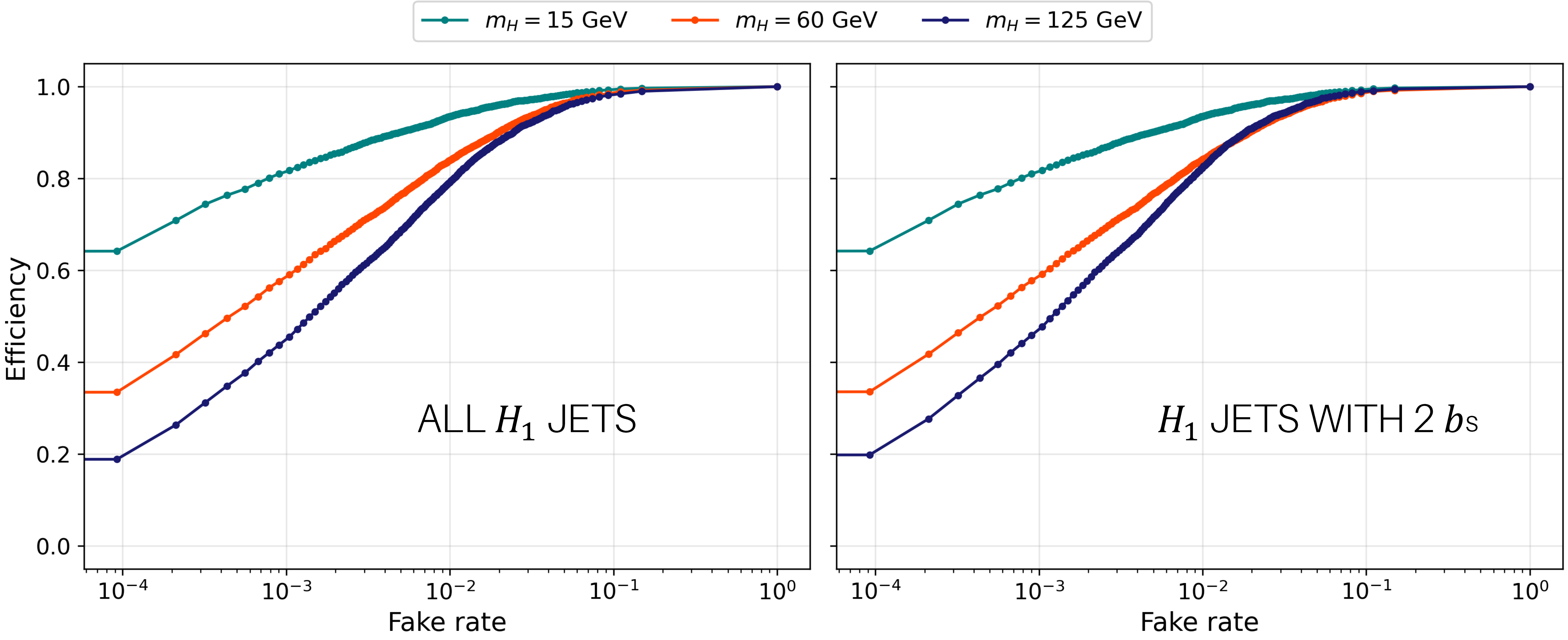
- Higher b-quark separation → Larger fraction of jets with one b-quark falling outside the cone → Reduced score



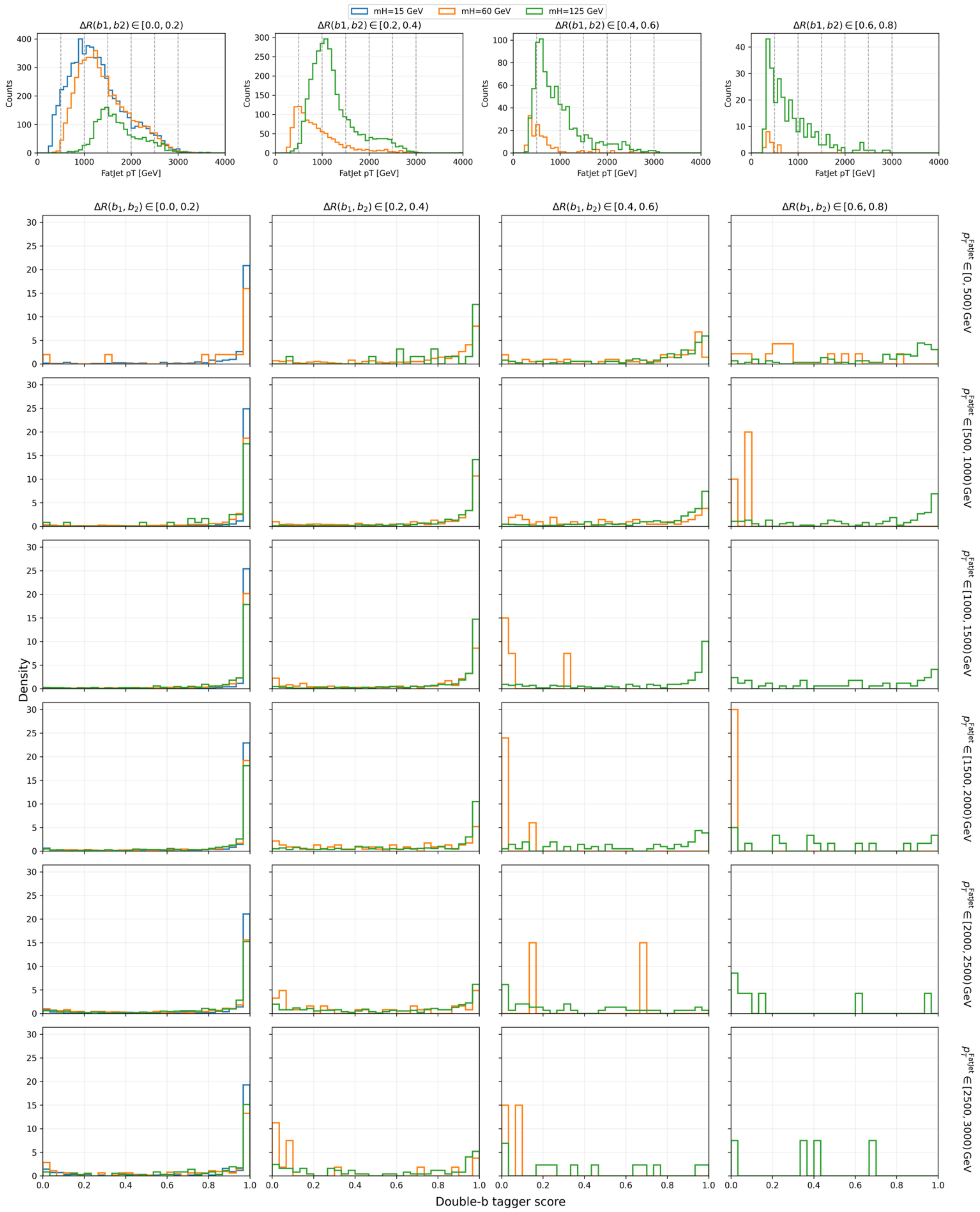


b-quark containment

- Higher b-quark separation → Larger fraction of fatjets with one b-quark falling outside the cone → Reduced score



ΔR_{bb} PERFORMANCE



ΔR_{bb} DEPENDENCE

- Naively expect large ΔR → more signal-like (QCD favours small, collinear splittings)
- But training is reweighted in m_{SD} and p_{T} → Compare at fixed kinematics
- Heuristic prediction based on 2-body dynamics & leading-order, soft collinear QCD
  - Signal peaked at $2m_{\mathrm{H}}/p_{\mathrm{T}}$, narrower distribution at small $2m_{\mathrm{H}}/p_{\mathrm{T}}$
  - QCD broad, falling as $\sim 1/\Delta R$ above kinematic minimum ($\Delta R \geq 2m/p_{\mathrm{T}}$)
- Implication:
  - Small $2m/p_{\mathrm{T}}$ → sharper signal peak vs broad QCD → better separation
  - Large $2m/p_{\mathrm{T}}$ → more overlap → reduced discrimination

A. Larkoski, An Unorthodox Introduction to QCD, arXiv:1709.06195 (2019) https://arxiv.org/pdf/1709.06195

J. Shelton, TASI Lectures on Jet Substructure, arXiv:1302.0260 (2013) https://arxiv.org/pdf/1302.0260

$\Delta R_{bb} \sim 2m_{\mathrm{H}}/p_{\mathrm{T}}$

Boosted 2-body decay approx. $H_1 \to b\bar{b}$

invariant mass of two ~massless particles:

$$m_H^2 = (p_b + p_{\bar{b}})^2 \quad \leftarrow m_b \ll p_{T,H}$$

$$= p_b^2 + p_{\bar{b}}^2 + 2p_b \cdot p_{\bar{b}}$$

4-mom. of massless particle: $p^\mu = (p_T \cosh\eta,\ p_T \cos\phi,\ p_T \sin\phi,\ p_T \sinh\eta)$

$$p_1 \cdot p_2 = p_{T_1} p_{T_2} (\cosh\eta_1 \cosh\eta_2 - \cos\phi_1 \cos\phi_2 - \sin\phi_1 \sin\phi_2 - \sinh\eta_1 \sinh\eta_2)$$

$$= p_{T_1} p_{T_2} (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))$$

$$= p_{T_1} p_{T_2} (\cosh\Delta\eta - \cos\Delta\phi)$$

$$m_H^2 = 2p_{Tb}\, p_{T\bar{b}} (\cosh\Delta\eta - \cos\Delta\phi)$$

Boosted decay → 2 quarks are close together → $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$ is small

small angle approxim: $\cosh\Delta\eta \sim 1 + \frac{\Delta\eta^2}{2}$

$$\cos\Delta\phi = 1 - \frac{\Delta\phi^2}{2}$$

$$\cosh\Delta\eta - \cos\Delta\phi \approx \frac{\Delta\eta^2 + \Delta\phi^2}{2} = \frac{\Delta R^2}{2}$$

$$m_H^2 = p_{Tb}\, p_{T\bar{b}}\, \Delta R$$

Assume Higgs decay ~ symmetric

$$p_{Tb} \sim p_{T\bar{b}} \sim \frac{p_T}{2}$$

$$m_H^2 \sim \left(\frac{p_T}{2}\right)^2 \Delta R^2 \Rightarrow m_H \sim \frac{p_T}{2} \Delta R$$

$\Delta R_{bb} \sim \frac{2m_H}{p_T}$ Distribution peaked at this value

small m_H → allowed ΔR range compressed to smaller values (narrower)

SMALL ΔR_{bb}

At signal peak, each B hadron carries transverse mom. $p_{T,B} \sim p_T/2$ for symmetric $z = 1/2$

Mean transverse decay length: $L_{xy} = \frac{p_{T,B}}{m_B} c\tau_B = \frac{p_T}{2m_B} c\tau_B$

Two B hadrons travel at $\pm \Delta R/2$ from jet axis → SVs separated in transverse plane by

$\Delta_{SV} \simeq L_{xy} \Delta R = \frac{p_T}{2m_B} c\tau_B \times \frac{2m}{p_T} = \frac{m c\tau_B}{m_B}$ → Transverse separation between SVs indep. of ΔR (and p_T), depending only on resonance mass

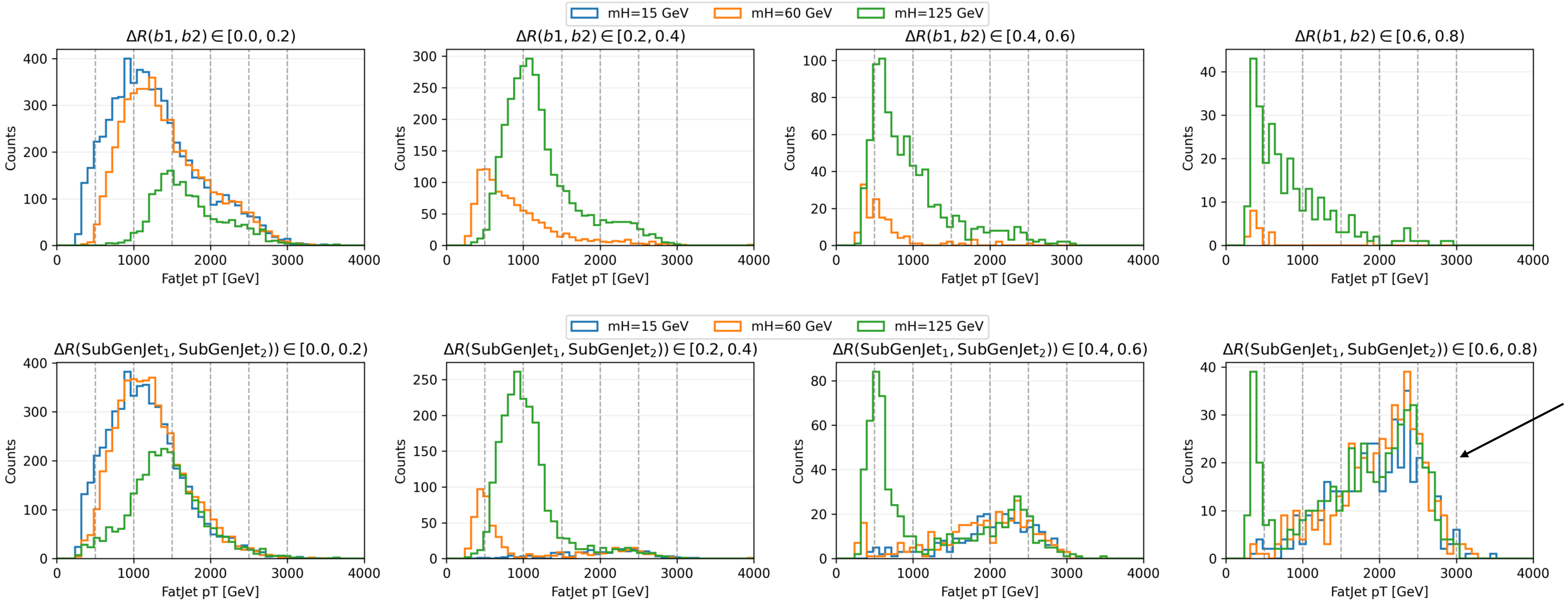
$m_B \approx 5.28$ GeV, $c\tau_B \approx 450\ \mu m$

$\Delta_{SV} \simeq 85 \times m\,[\mathrm{GeV}]\ \mu m$

CMS SV resol. $\sim 200\ \mu m$ → SVs resolvable as long as $m \gtrsim 2.4$ GeV

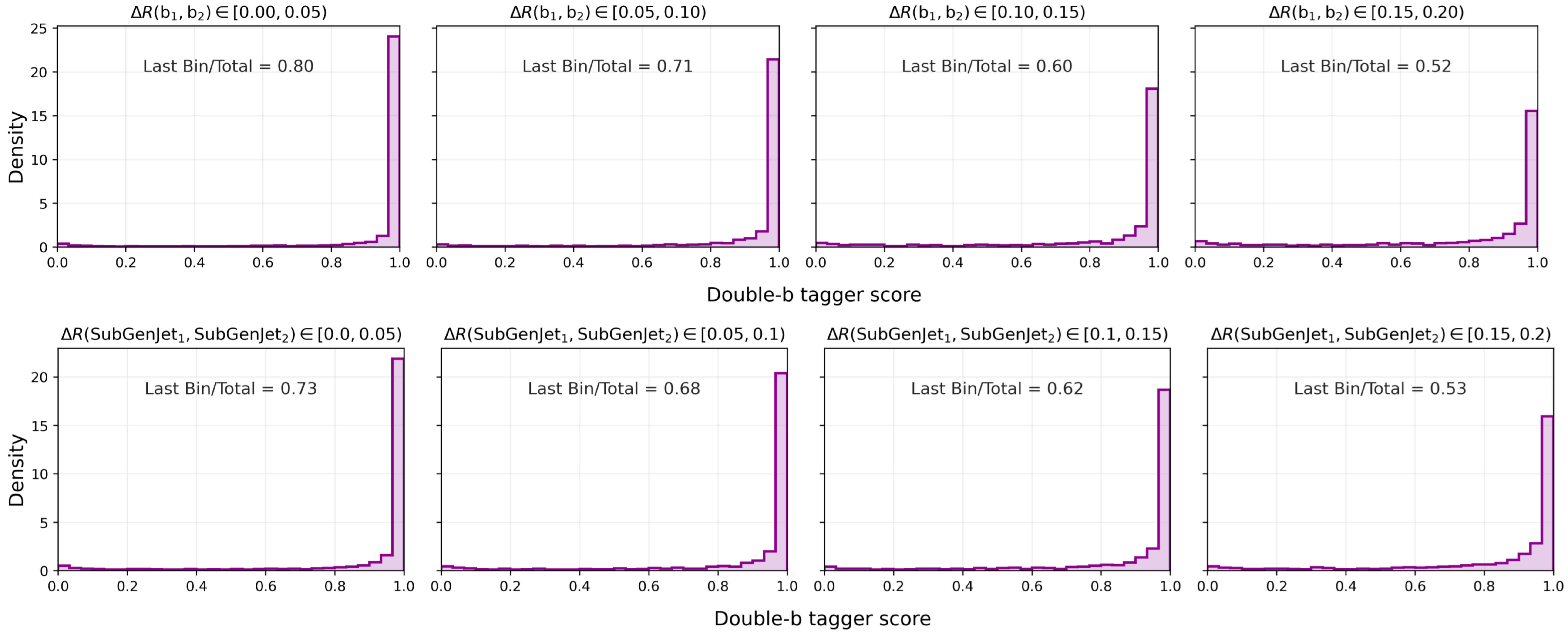
ΔR_{bb} PROXY VARIABLE STUDIES

p_T distributions for fatjets in ΔR_{bb} bins



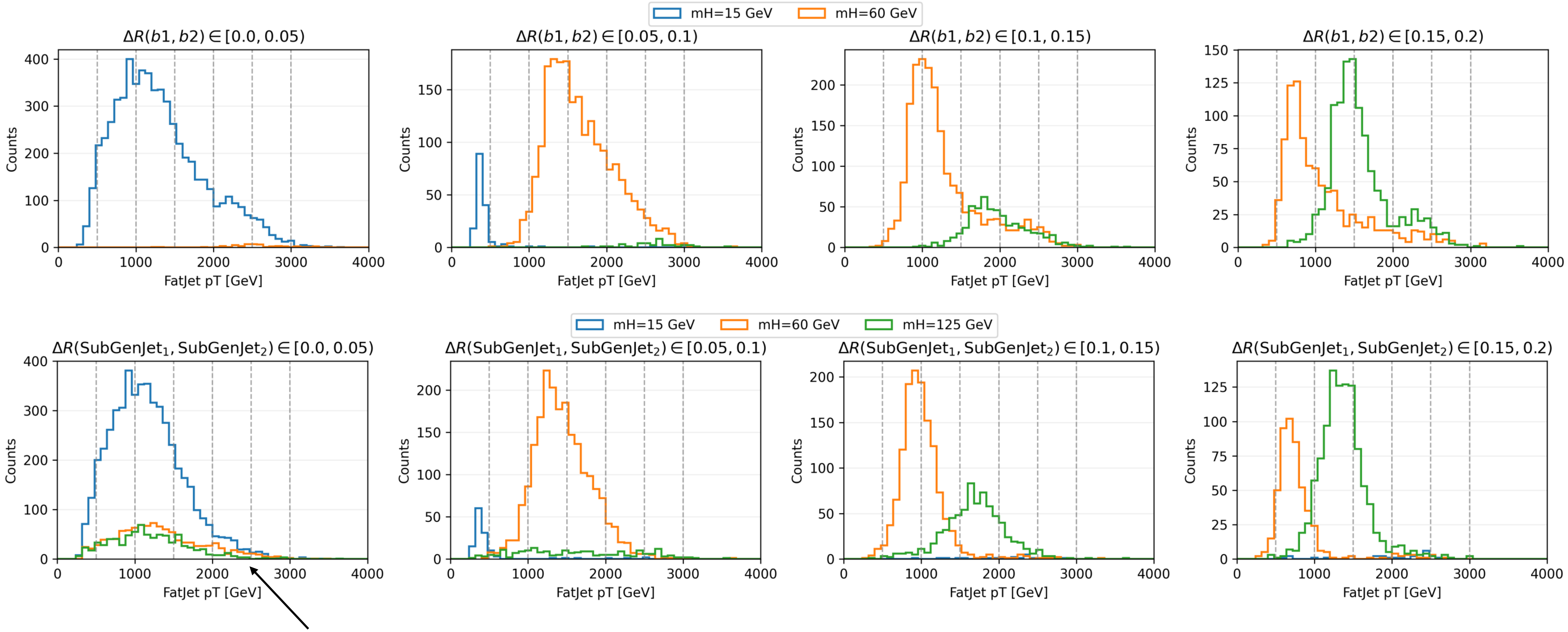
Migration to higher ΔR_{bb} bins is largely due to two b-quarks matched to the same SubGenJet

ΔR_{bb} PROXY VARIABLE STUDIES



ΔR_{bb} PROXY VARIABLE STUDIES

p_{T} distributions for fatjets in ΔR_{bb} bins (small range)



Migration to lowest ΔR_{bb} bin not understood

QCD m_{SD} BINS