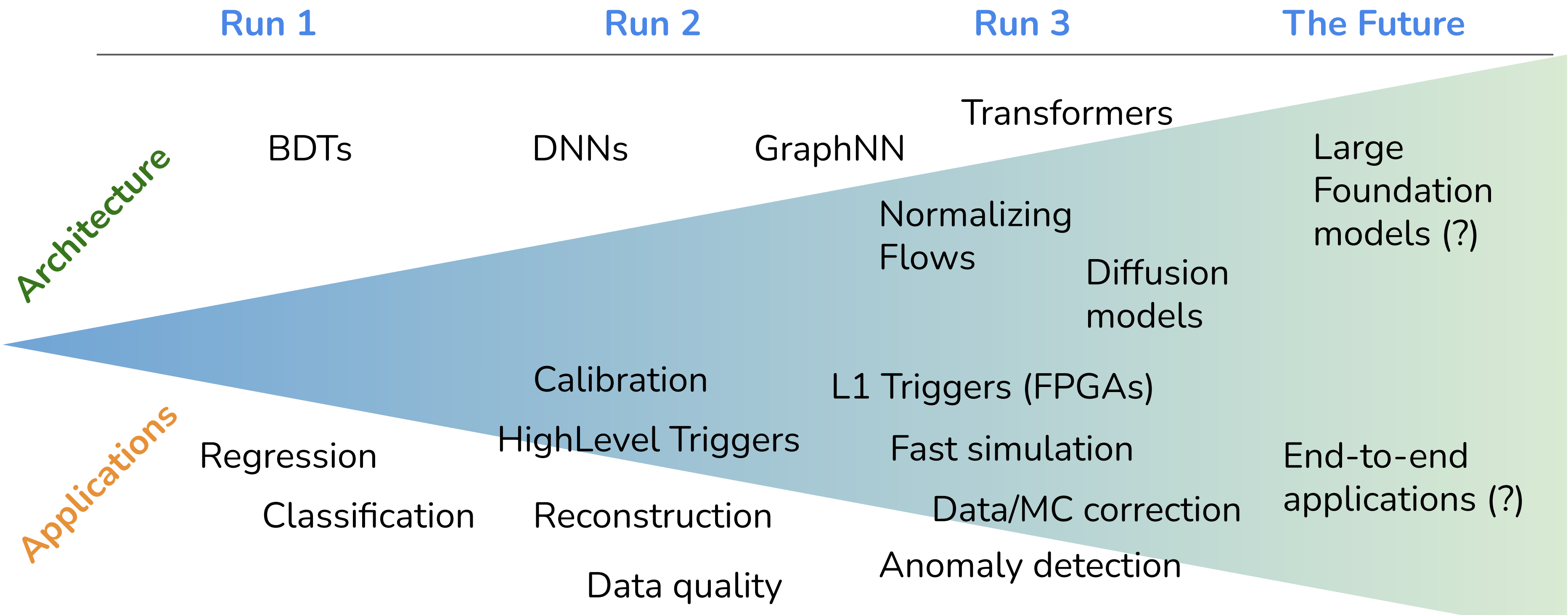

Davide Valsecchi (ETH Zurich) - CMS
Johannes Michael Wagner (University of California Berkeley) - ATLAS

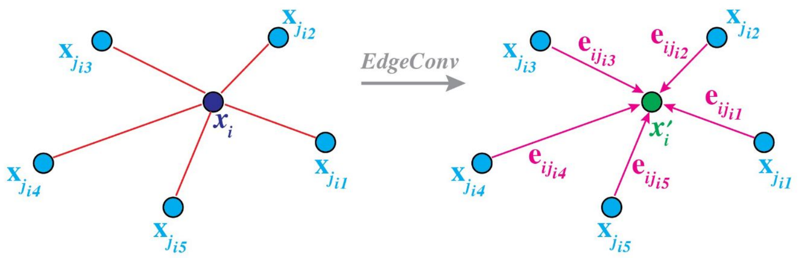
# Motivation

ML expanding toolbox → many interesting applications in HEP beyond Sig vs Bkg



# ML architectures used in HEP

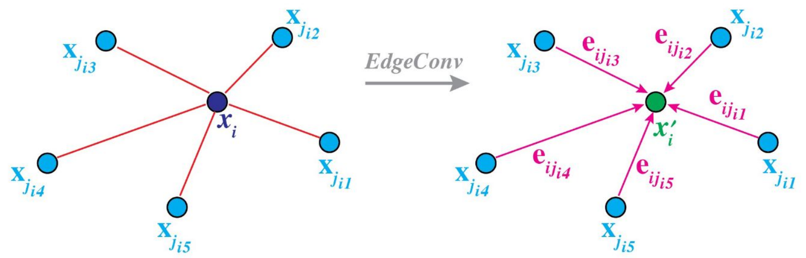
ML architecture depends on the data structure and task



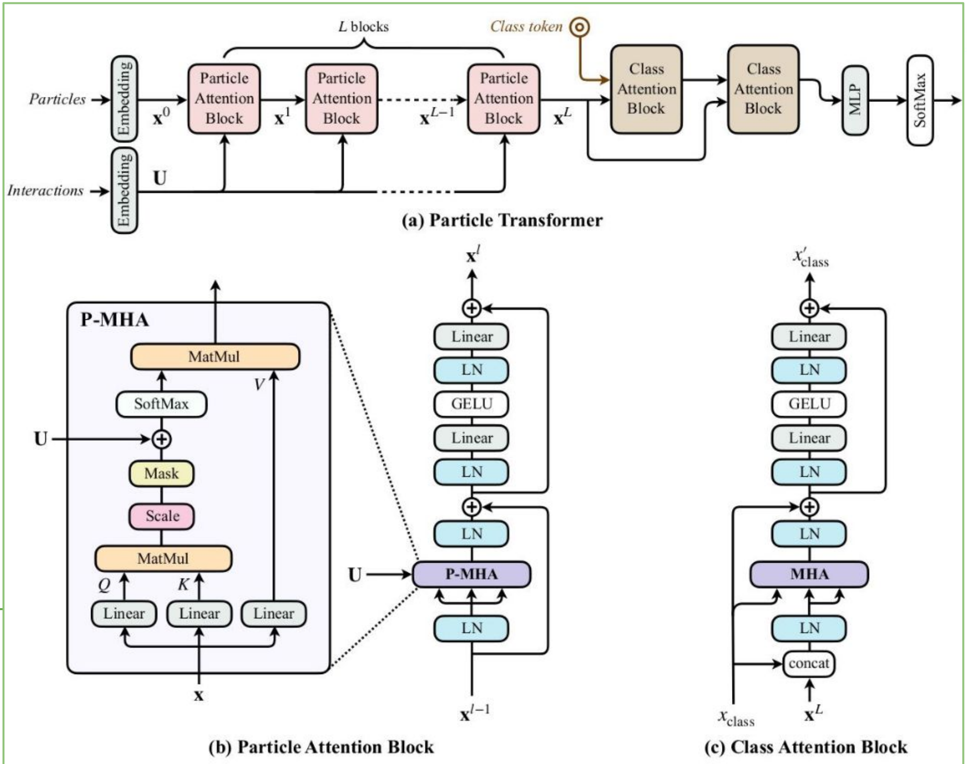

**Graph Networks**: encode the relation between (unordered) objects. Extract information both about the full graph and the single objects or edges. [arxiv1812.08434](https://arxiv.org/abs/1812.08434)

# ML architectures used in HEP

ML architecture depends on the data structure and task




**Graph Networks**: encode the relation between (unordered) objects. Extract information both about the full graph and the single objects or edges.
arxiv1812.08434



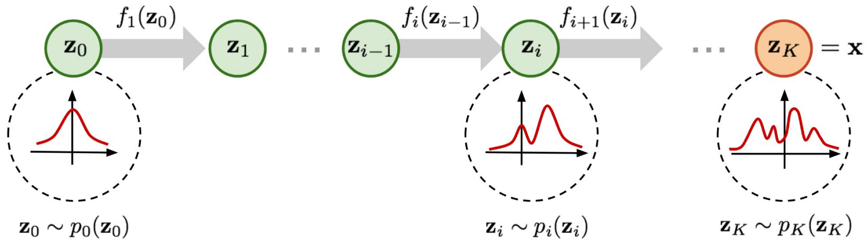

**Transformer**: like fully-connected graph network, but introduces an efficient **attention mechanism.**

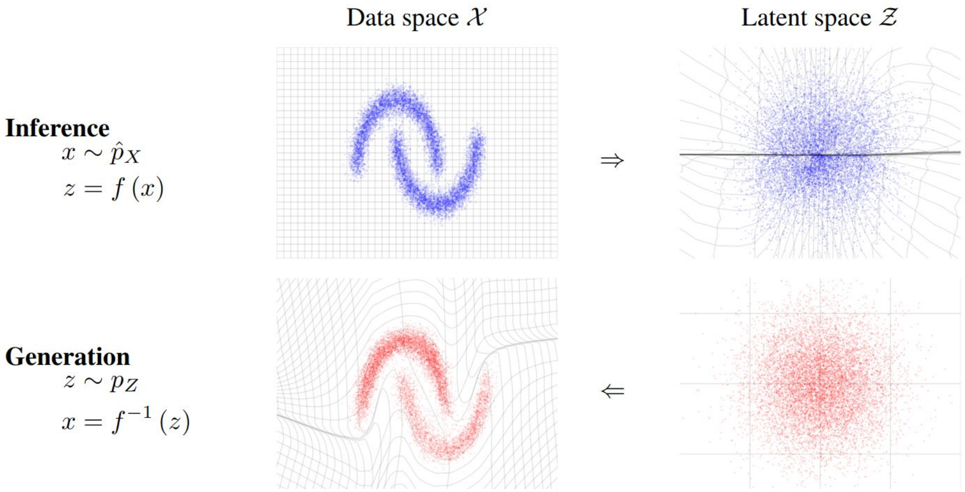
State-of-the-art (SOTA) architecture for many HEP tasks.

**Particle Transformer** → inputs == particles → added meaningful pairwise features ($k_T$, $m_{inv}$ ..) in the attention matrix
arxiv1706.03762, arxiv2106.03898, arxiv2202.03772

# ML architectures used in HEP

ML architecture depends on the data structure and task




Probabilistic ML architectures:
**Normalizing Flows** arxiv1908.09257

→ Used to model complex p.d.f. with a sequence of simple invertible transformations from a simple base p.d.f. (gaussians)

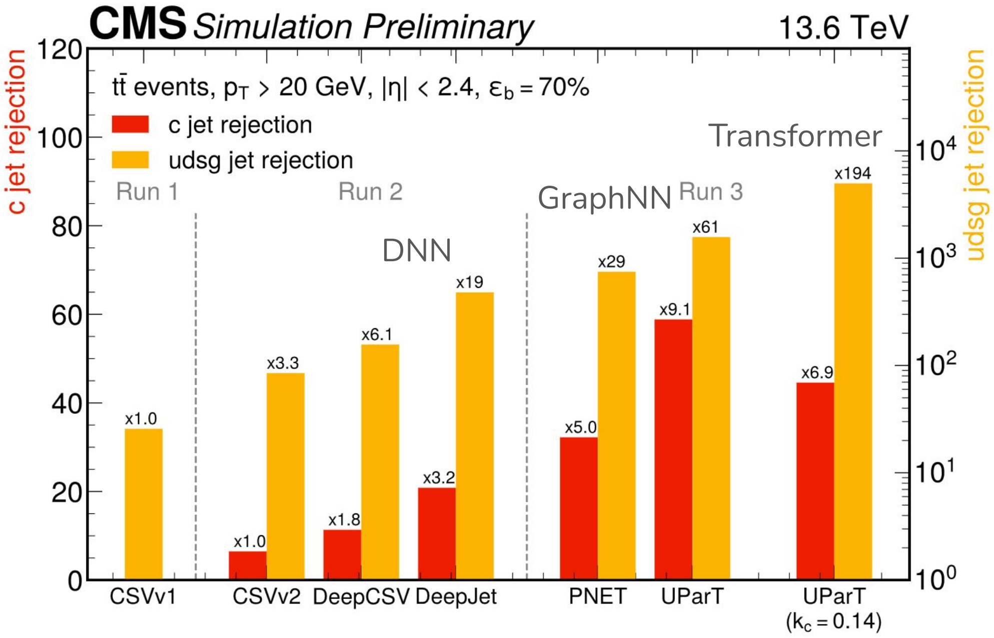
→ both **density** estimation and sample **generation**

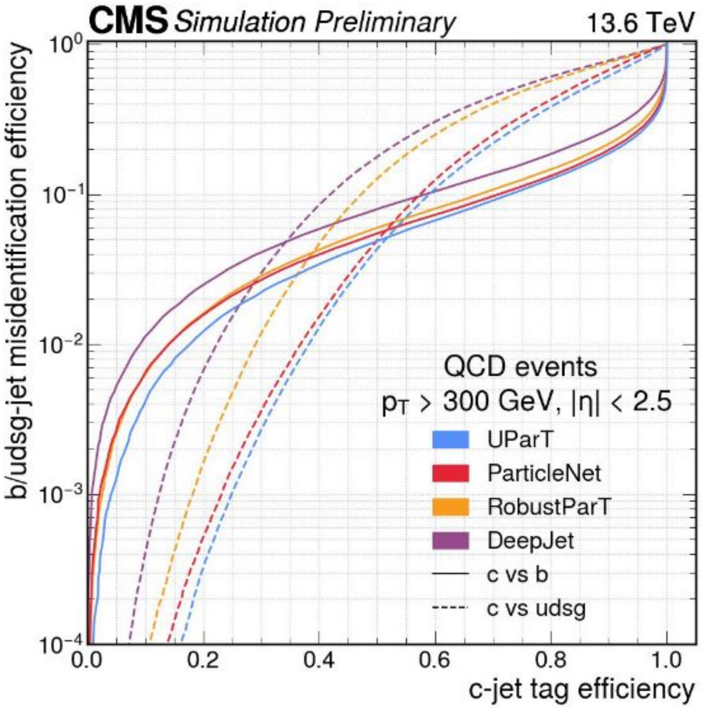
Many natural applications in HEP:
- fast simulation: CMS FlashSim CMS-DP-2024-080
- calibration (CMS-PAS-HIG-23-014, ATLAS-CONF-2024-014)
- importance sampling CMS-DPS-2023-085v2

# A unified approach to Jet Tagging

Evolution of taggers → From **DNN** to **GraphNN** (ParticleNet) to **Transformer** (UnifiedParT)

- Many improvements all included in the new state-of-the-art tagger CMS-DP-2024-066
- “Unified” tagger → **UParT**: new Run3 best model

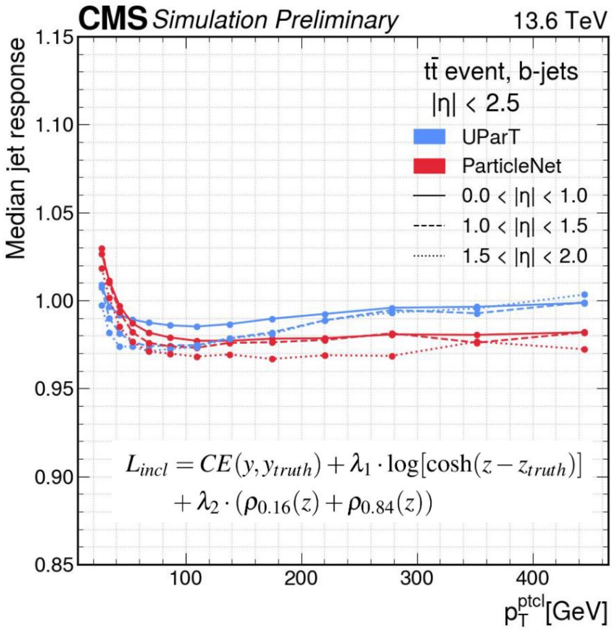


# UParT: robust training and jet calibration

CMS-DP-2024-066

- Training includes also jet $p_T$ regression and resolution estimation



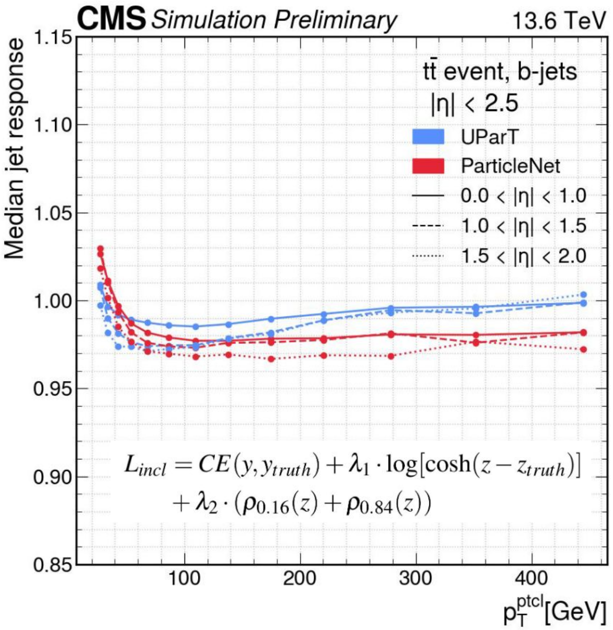

UPart has a better jet response, especially at high eta

# UParT: robust training and jet calibration

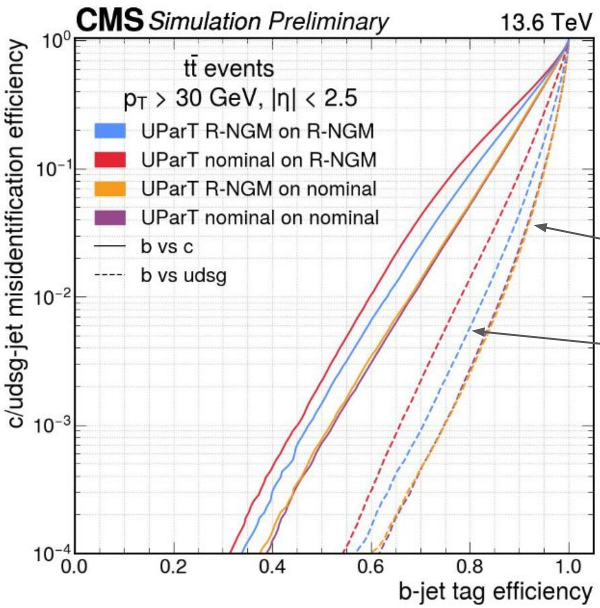
CMS-DP-2024-066

ETH zürich

- Training includes also jet $p_T$ regression and resolution estimation
- Trained with adversarial attack for data/MC robustness




UPart has a better jet response, especially at high eta




Training with adversarial attack have **same** performance on nominal and **better** on varied data.

Rectified Normed Gradient Method (R-NGM) arxiv1901.08573

$$x_{i,adv} = x_i + \epsilon \cdot |\nabla CE(x_i, \theta)|$$

$$L = CE(x_{i,adv}, \theta) + \lambda \cdot KL(x_i, x_{i,adv})$$

Train the model on distorted inputs to become less sensitive to feature changes

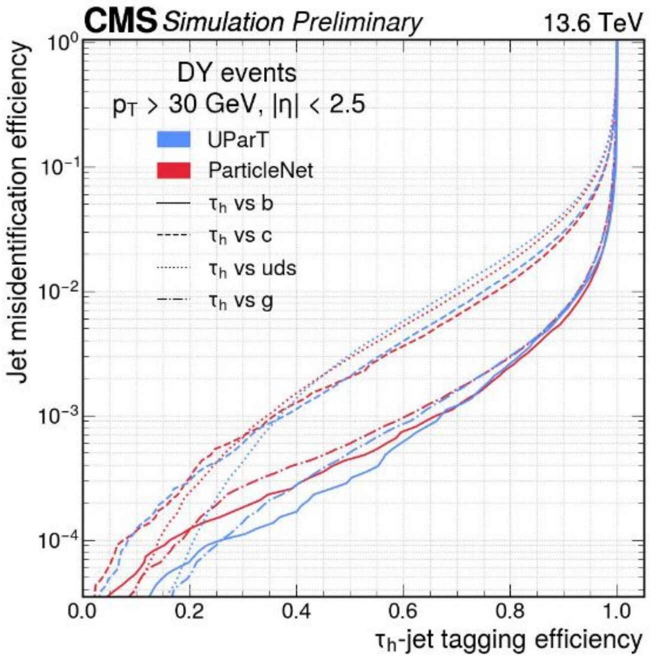
## UParT: s-tagging and tau-tagging

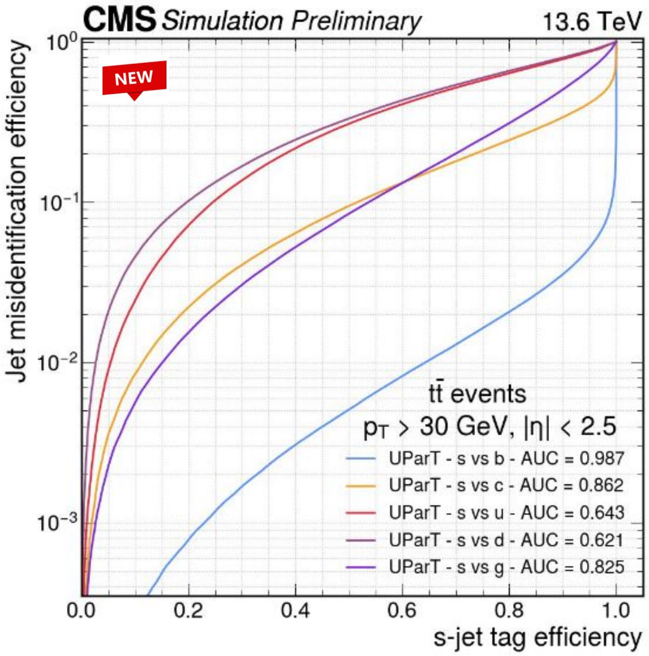


**UPart** includes classes for **hadronic taus**, and **strange** jets.

- made possible by new tuning of CMS pileup removal algorithm CMS-DP-2024-043

- low-efficiency s-tagger: **first time** in CMS

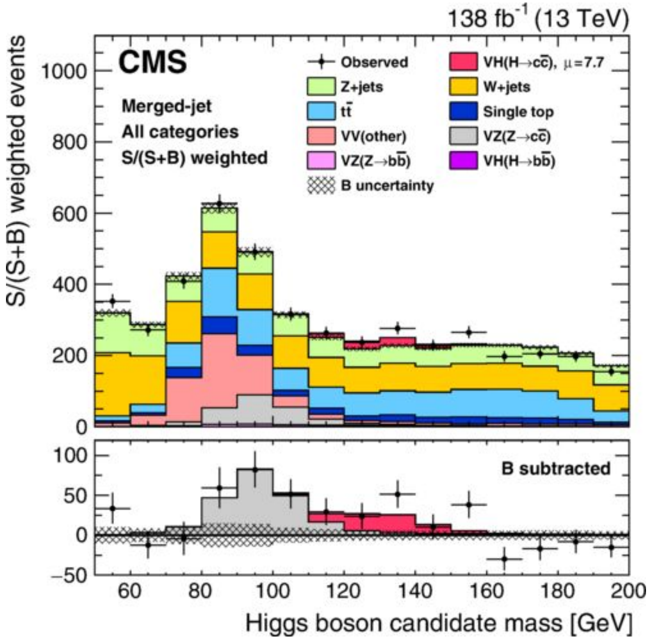


# Boosted Jet taggers

**ParticleNet** is very powerful also for searching for resonances in AK8 jets.

- Implemented in Run2 for boosted X(bb) and X(cc) tagging
- significant improvements in the performance of VH(cc) PhysRevLett.131.061801 and HH(4b) PhysRevLett.131.041803 searches (see talk from Raffaele)

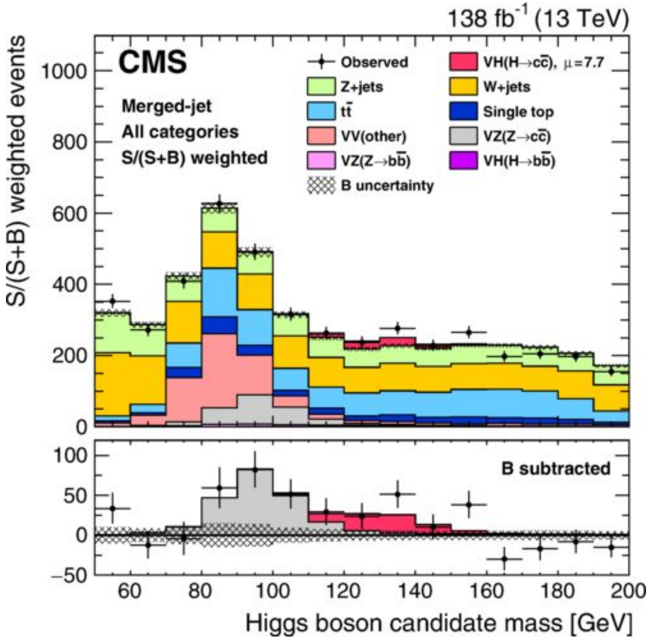



VH(cc) PhysRevLett.131.061801

# Boosted Jet taggers

**ParticleNet** is very powerful also for searching for resonances in AK8 jets.

- Implemented in Run2 for boosted X(bb) and X(cc) tagging
- significant improvements in the performance of VH(cc) PhysRevLett.131.061801 and HH(4b) PhysRevLett.131.041803 searches (see talk from Raffaele)

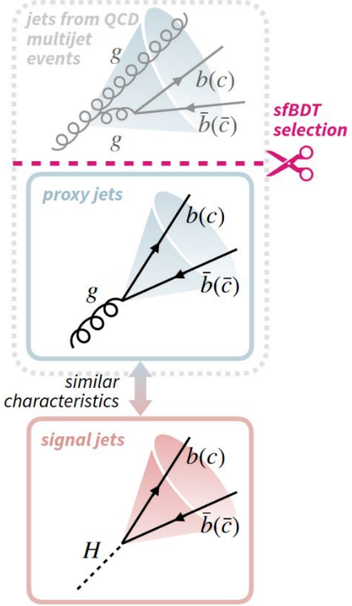



VH(cc) PhysRevLett.131.061801

3 calibration methods documented in BTV-22-001

Using ML for calibration: **sfBDT**

- Select g→bb/cc jets in QCD that are more similar to X→bb/cc decays
- Inputs: N-subjettines, track and SV information



# Boosted Jet taggers

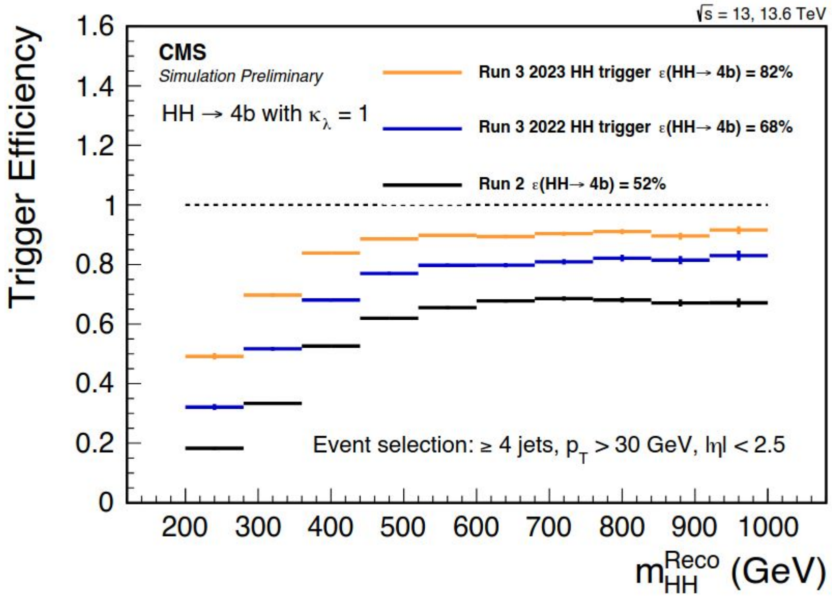
Many improvements for Run3: CMS-DP-2024-055

- added boosted H($\tau\tau$) decays
- Simultaneous train classification and regression tasks in single network
- Improve mass decorrelation by sampling more granularly the mX range: no mass sculpting on background CMS-DP-2021-017

From Run3 using ParticleNet-MD in **scouting trigger**s to enhance the acceptance of analysis looking at final state with boosted hadronic resonances.

More info in M. Stamenkovic talk at FTAG workshop

CMS-DP-2023-050



Exploring an extension of PartT to multi-prong decays in AK8 jets → **GloParT**

- developed for HH→VV→4b analysis [CMS-PAS-HIG-23-012]
- including: top-tagging (3 prong), X → VV with hadronic and leptonic decays, taus etc.
- same mass decorrelation strategy by using flat mass signal samples vs QCD → learn only substructure

**Challenge**: calibration with multi-prong decays

| Process | Final state/ prongness | heavy flavour | # of classes |
|---|---|---|---|
| H→VV (full-hadronic) | qqqq | 0c/1c/2c | 3 |
| | qqq | | 3 |
| H→WW (semi-leptonic) | eνqq | 0c/1c | 2 |
| | μνqq | | 2 |
| | $\tau_e\nu qq$ | | 2 |
| | $\tau_\mu\nu qq$ | | 2 |
| | $\tau_h\nu qq$ | | 2 |
| H→qq | | bb | 1 |
| | | cc | 1 |
| | | ss | 1 |
| | | qq (q=u/d) | 1 |
| H→ττ | $\tau_e\tau_h$ | | 1 |
| | $\tau_\mu\tau_h$ | | 1 |
| | $\tau_h\tau_h$ | | 1 |
| t→bW (hadronic) | bqq | 1b + 0c/1c | 2 |
| | bq | | 2 |
| t→bW (leptonic) | beν | 1b | 1 |
| | bμν | | 1 |
| | $b\tau_e\nu$ | | 1 |
| | $b\tau_\mu\nu$ | | 1 |
| | $b\tau_h\nu$ | | 1 |
| QCD | | b | 1 |
| | | bb | 1 |
| | | c | 1 |
| | | cc | 1 |
| | | others (light) | 1 |

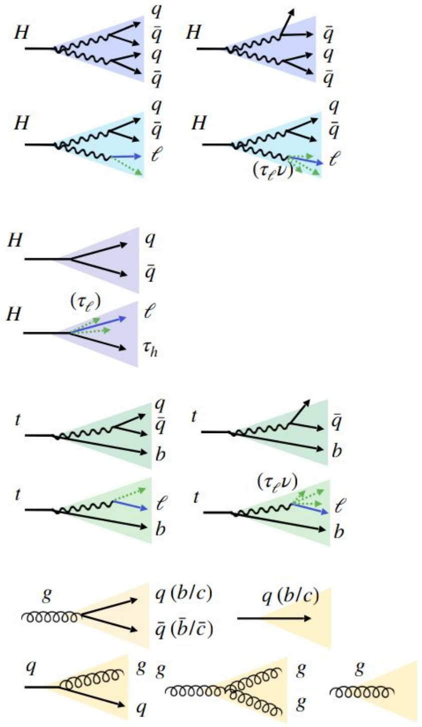
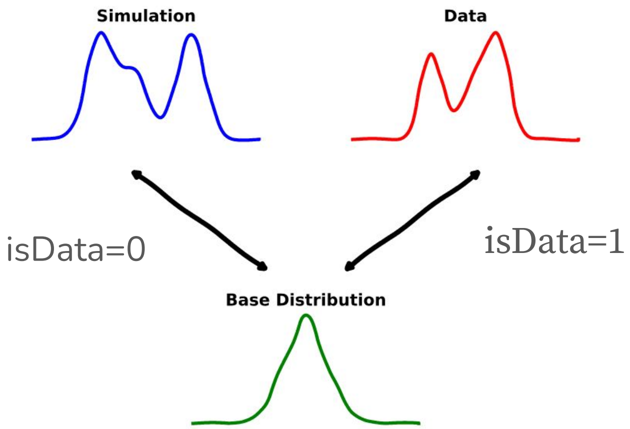



Figure 4: Full set of training jet classes for GloParT.




**Problem**: correct MC mis-modeling of complex IDs or regressions depending on many observables.

**Normalizing Flows (NF)** can be used to build an “optimal transport” correction:

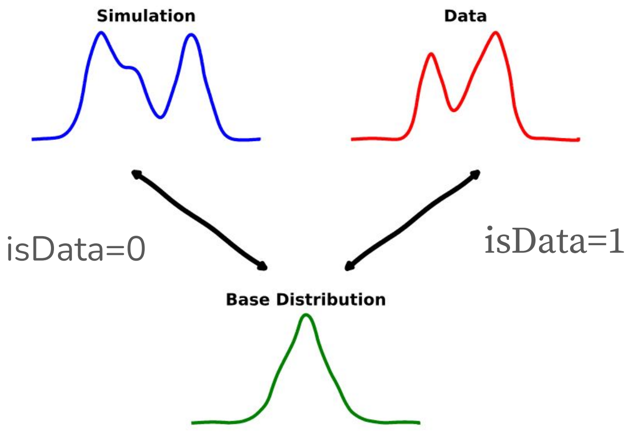
→ train a NF to bring data and MC p.d.f.s to a common base distribution (gaussian)

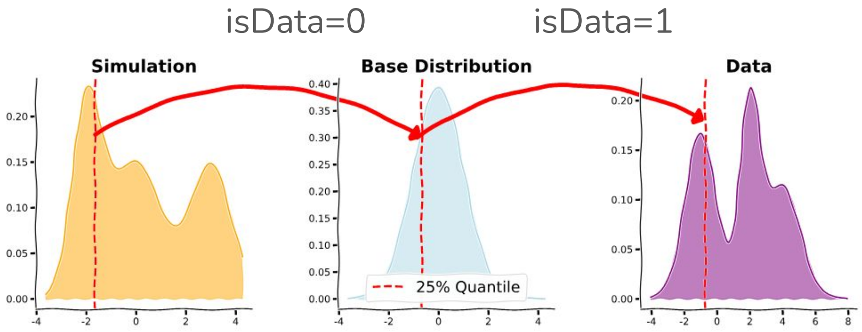
→ use the resulting transformations to morph MC into data

→ the transformation can be made **conditional** on other observables

Demonstrated method on toy data:
*One Flow to correct them all* arxiv2403.18582

# MC-to-Data calibration with Normalizing Flows




**Problem**: correct MC mis-modeling of complex IDs or regressions depending on many observables.

**Normalizing Flows (NF)** can be used to build an “optimal transport” correction:

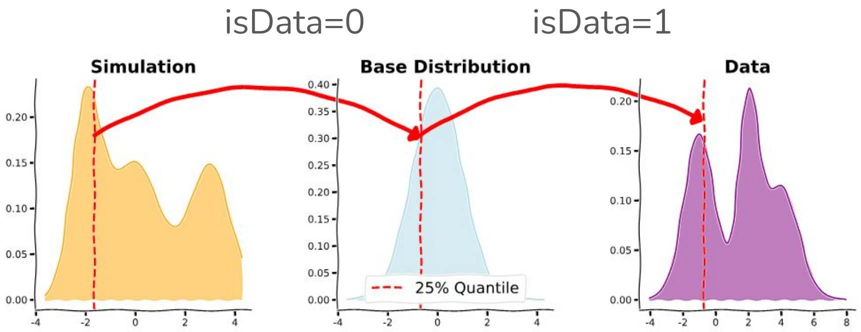
→ train a NF to bring data and MC p.d.f.s to a common base distribution (gaussian)

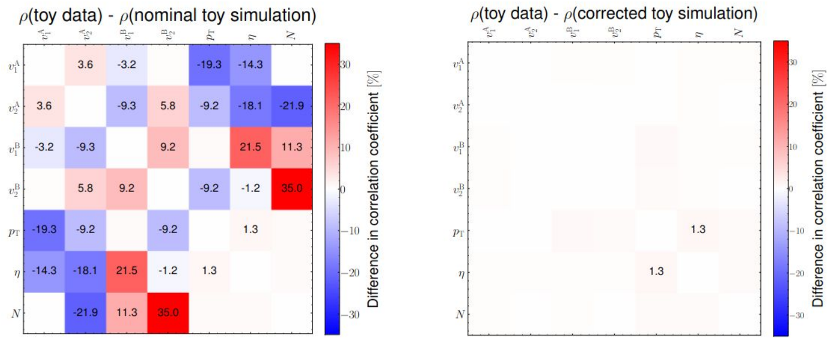
→ use the resulting transformations to morph MC into data

→ the transformation can be made **conditional** on other observables

Demonstrated method on toy data:
*One Flow to correct them all* arxiv2403.18582

# MC-to-Data calibration with Normalizing Flows




Correct both 1D distributions and (also complex) correlations

**Problem**: correct MC mis-modeling of complex IDs or regressions depending on many observables.

**Normalizing Flows (NF)** can be used to build an "optimal transport" correction:

→ train a NF to bring data and MC p.d.f.s to a common base distribution (gaussian)

→ use the resulting transformations to morph MC into data

→ the transformation can be made **conditional** on other observables

Demonstrated method on toy data:
*One Flow to correct them all* [arxiv2403.18582](arxiv2403.18582)

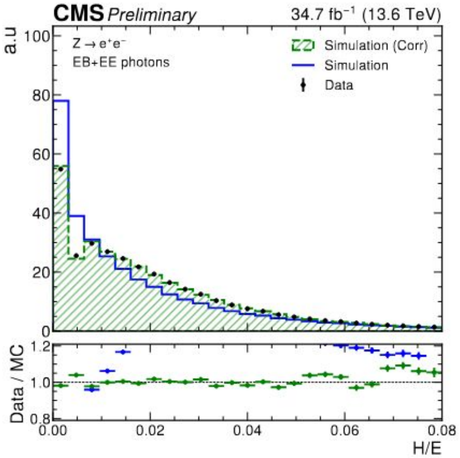
# MC-to-Data calibration with Normalizing Flows

H → γγ inclusive and differential XS at 13.6 TeV (Run3 2022 data)
CMS-PAS-HIG-23-014

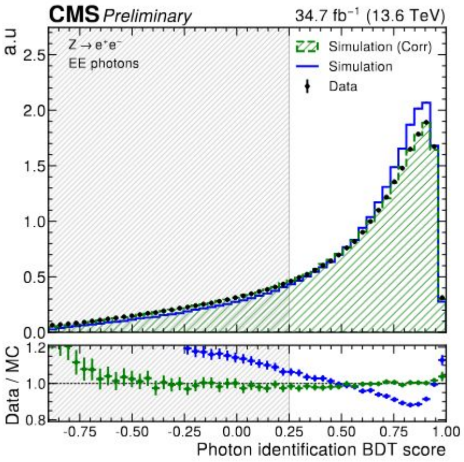
**Photon ID inputs**: shower shapes, isolation variables, H/E, per-photon energy resolution estimate $\sigma_E$

Mismodeling Photon ID → mismodeling of the per-event diphoton inv-mass resolution $\sigma_M$ → affects analysis categorization → **large uncertainty**

- In Run2 used a **chain of BDT** to correct shower shape and isolation variables
- in Run3 used the **Normalizing Flow** technique to correct shower shape, isolation and $\sigma_E$, conditionally by η, $\varphi$, pT of the photon and ρ (energy density from PU)
    - **Simpler model and even more powerful correction**: ~1-2% residual data/MC discrepancy checked in Z → μμγ events




H/E = had./electr. reconstructed energy



# ML for EFT

We can exploit the EFT quadratic expansion to build optimal classifiers to measure SMEFT operators effects.

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int p(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} = \int p(\boldsymbol{x}|\boldsymbol{z}) p(\boldsymbol{z}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{z},$$

the full likelihood is intractable at detector level (x).
MELA method attacks this with transfer functions

likelihood ratio

$$R(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \frac{\mathrm{d}\sigma_{\boldsymbol{\theta}}(\boldsymbol{x})/\mathrm{d}\boldsymbol{x}}{\mathrm{d}\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{x})/\mathrm{d}\boldsymbol{x}} = \frac{\sigma(\boldsymbol{\theta})\, p(\boldsymbol{x}|\boldsymbol{\theta})}{\sigma(\boldsymbol{\theta}_0)\, p(\boldsymbol{x}|\boldsymbol{\theta}_0)} \longrightarrow r(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\theta}_0) \equiv \frac{p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{p(\boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{z}|\boldsymbol{\theta}_0)}$$

joint likelihood ratio: reco, hadronization, showering factors cancel

We can exploit the EFT quadratic expansion to build optimal classifiers to measure SMEFT operators effects.

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \int p(\boldsymbol{x},\boldsymbol{z})\mathrm{d}\boldsymbol{z} = \int p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{z},$$

the full likelihood is intractable at detector level (x). MELA method attacks this with transfer functions

likelihood ratio

$$R(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\theta}_0) = \frac{\mathrm{d}\sigma_{\boldsymbol{\theta}}(\boldsymbol{x})/\mathrm{d}\boldsymbol{x}}{\mathrm{d}\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{x})/\mathrm{d}\boldsymbol{x}} = \frac{\sigma(\boldsymbol{\theta})\,p(\boldsymbol{x}|\boldsymbol{\theta})}{\sigma(\boldsymbol{\theta}_0)\,p(\boldsymbol{x}|\boldsymbol{\theta}_0)} \longrightarrow r(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\theta},\boldsymbol{\theta}_0) \equiv \frac{p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{\theta}_0)} = \frac{p(\boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{z}|\boldsymbol{\theta}_0)}$$

joint likelihood ratio: reco, hadronization, showering factors cancel

$$\boxed{L = \sum_{\theta\in\mathcal{B}} \int \mathrm{d}x\,\mathrm{d}z\, p(x,z|\mathrm{SM}) \left( r(x,z|\theta)\hat{f}(x;\theta)^2 + (1-\hat{f}(x;\theta))^2 \right)} \longrightarrow$$

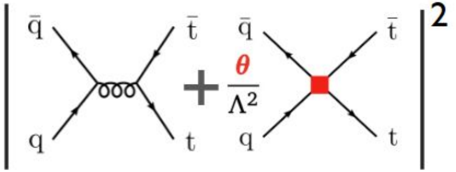
likelihood ratio trick

$$\hat{f}(x;\theta) = \frac{1}{1+\hat{R}(x;\theta)}$$

We can use a ML model to approximate f(X,θ) with a regression → optimal likelihood ratio for each θ point in the Wilson space

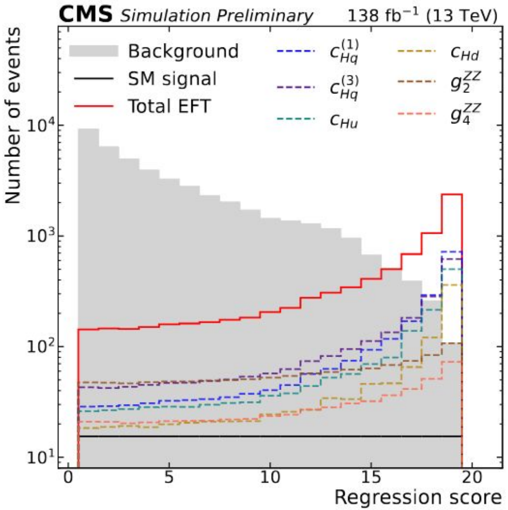
$$R(x|\theta,\theta_0) = 1 + \sum_a (\theta-\theta_0)_a R_a(x) + \sum_{a,b} \frac{1}{2}(\theta-\theta_0)_a(\theta-\theta_0)_b R_{ab}(x)$$

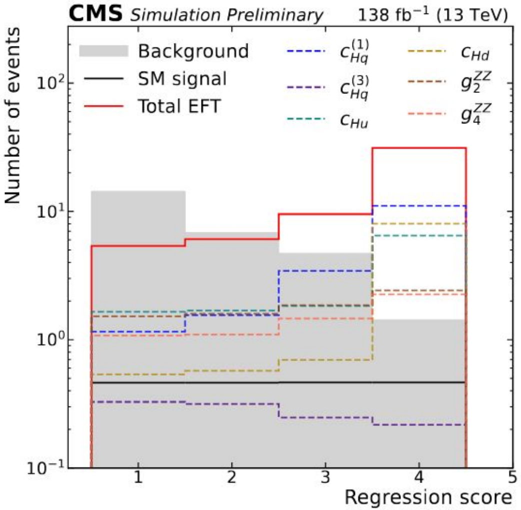
**polynomial expansion**: model each term with BDTs or NNs




More details in Learning EFT with Tree boosting arxvi2205.12976, MadMiner: 1907.10621

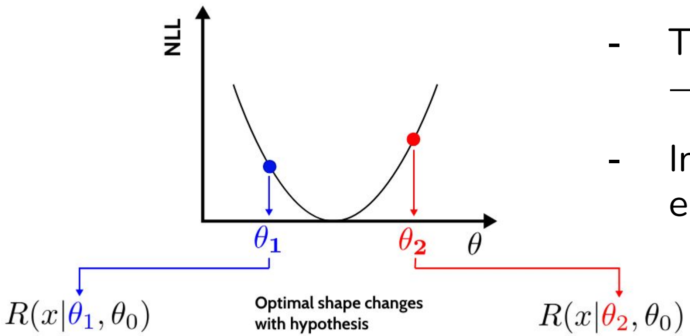
# Boosted Information Trees for EFT fits




VH→bb EFT interpretation
CMS-PAS-HIG-23-016

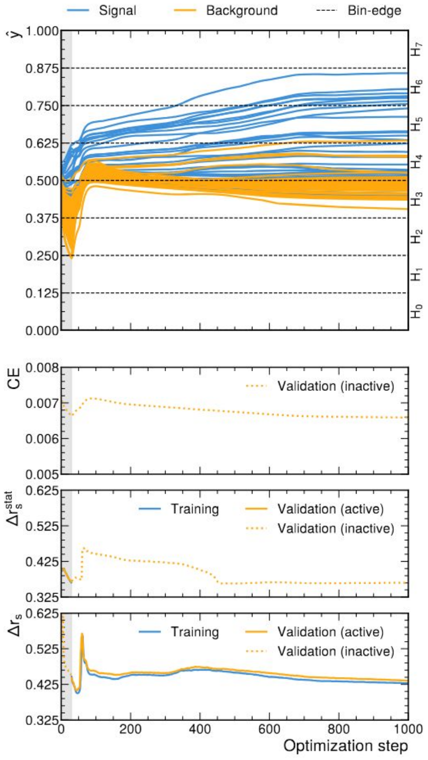
- Using Boosted Information Trees arxiv to build parametrized classifier using angular observables as inputs
- See talk on Wednesday from V. Perovic



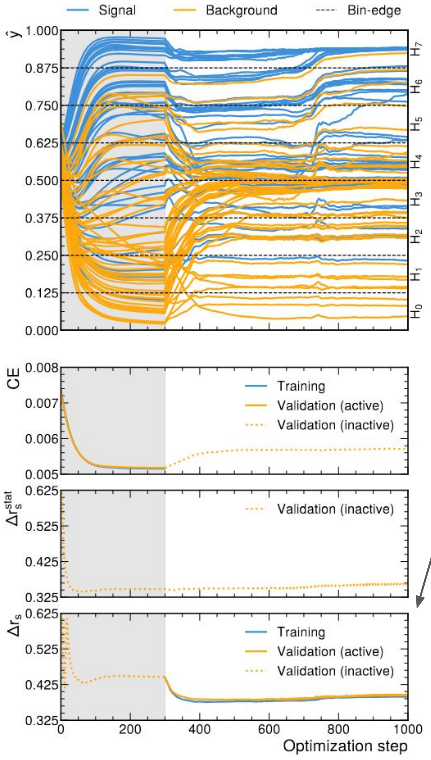

- The classifier is **optimal for each Wilson** coefficient point θ → binned analysis → **need to choose a single shape**
- Implemented optimization procedure to choose the best point that ensure sensitivity to **multiple Wilson coefficients**

# Systematic-aware NN trainings

Development of systematic-aware NN trainings for binned-likelihood-analyses at the LHC [CMS-PAS-MLG-23-005]




vanilla classifier

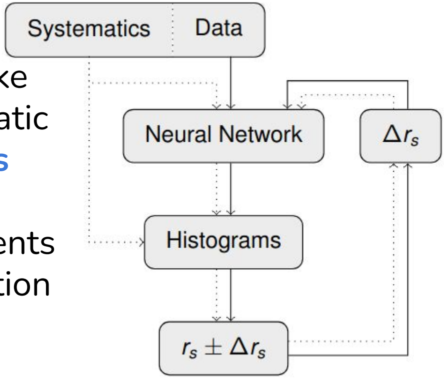


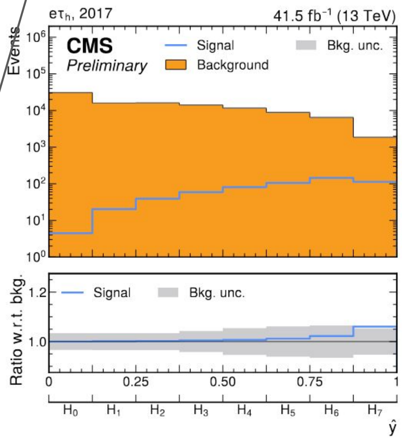

systematics aware training

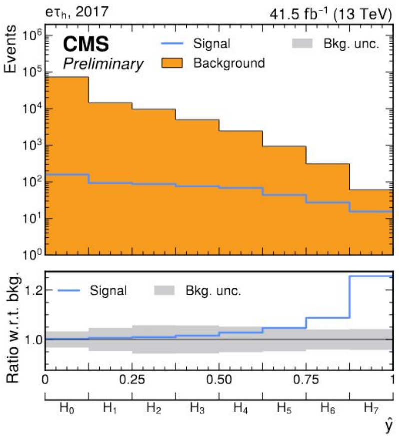
bins

**Special training procedure** to take into account the effect of systematic uncertainty on **final analysis bins**

→ need to **backpropagate** gradients including $\Delta r_s$ systematic information



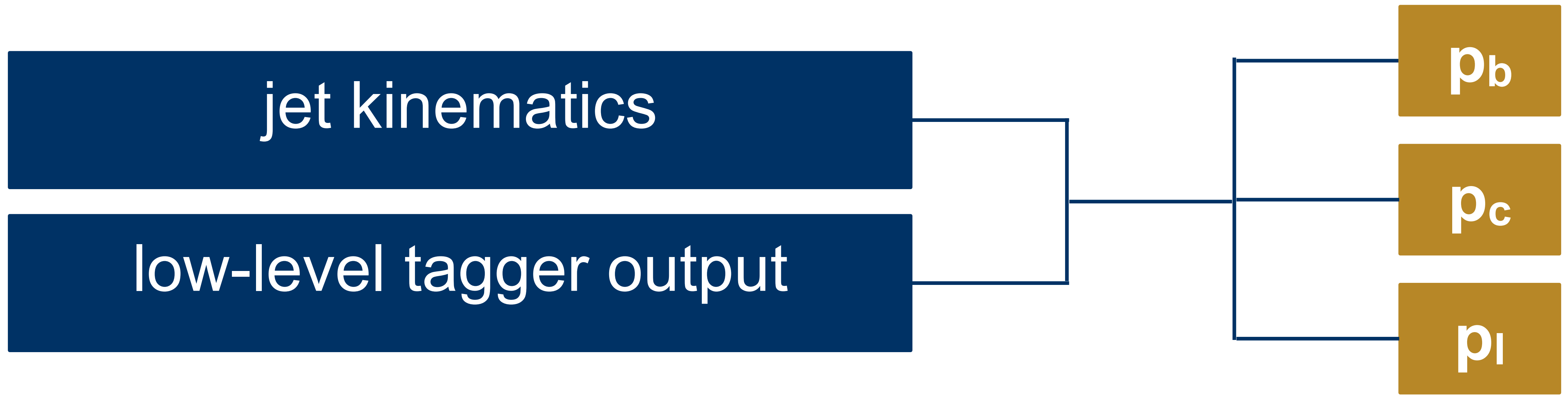



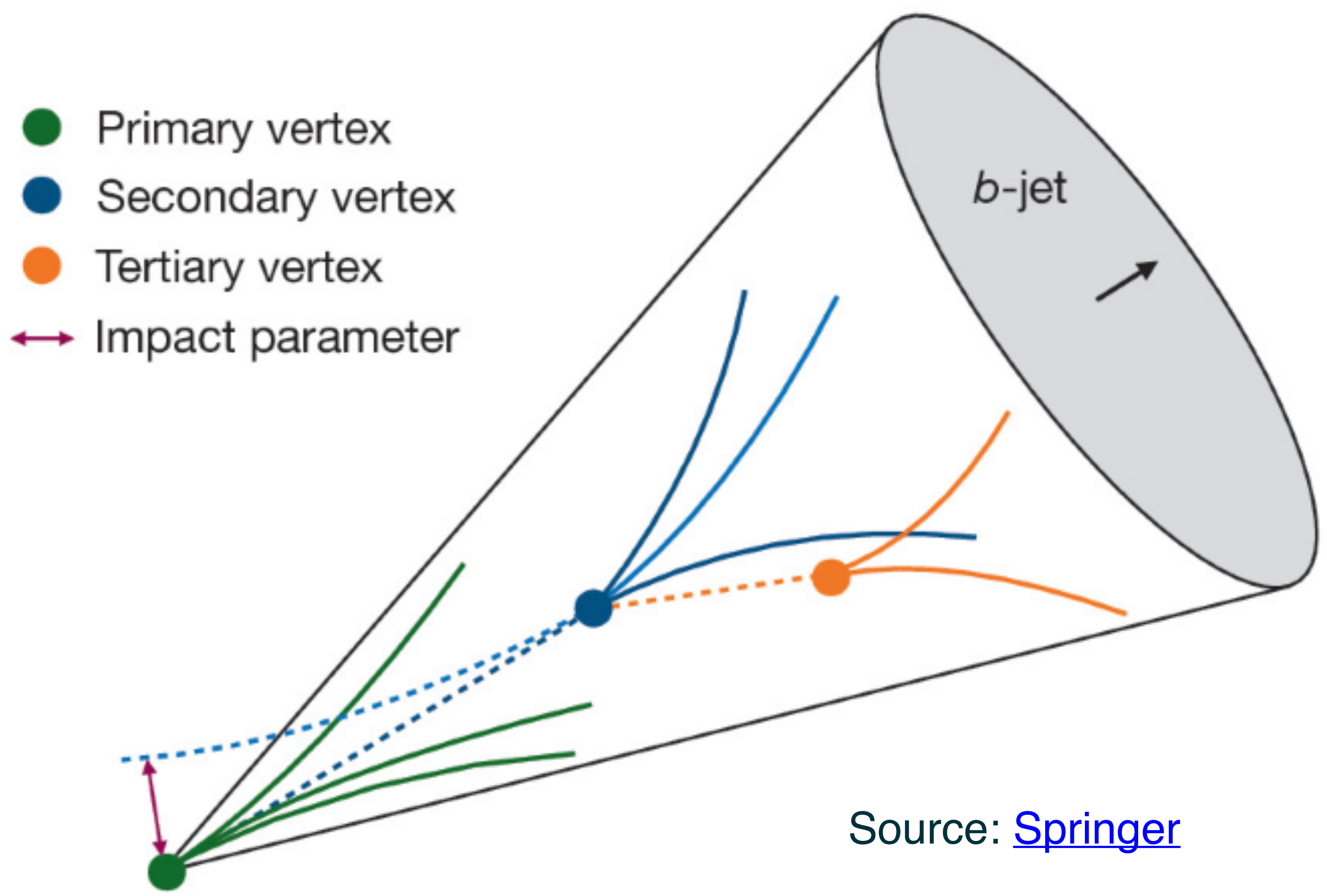

→ 12 and 16%, better sensitivity in $r_{ggH}$ and $r_{qqH}$ (STXS stage-0, eτ final state)

# Flavor tagging: run 2 improvements



- Many Higgs measurements rely on powerful **jet flavor tagging algorithms**
  - High $p_T$ (boosted) and low $p_T$ (resolved) topology require **different approaches**
- Run 2 standard for resolved flavor tagging is **DNN-based DL1r** [FTAG-2019-07]



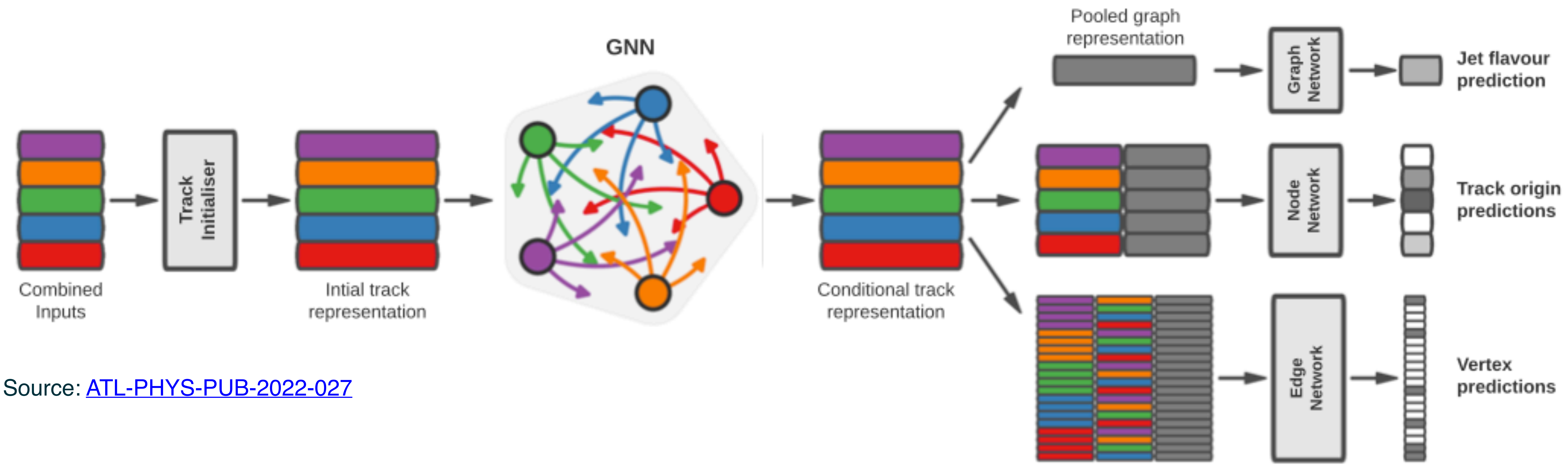

Source: Springer

  - Low level taggers consist of **IP** and **SV** fitting algorithms plus **track-based RNN**
- Separate DNN-based algorithm in boosted regime **uses DL1r scores on VR track jets** to classify physics of large-R jets as **H(bb), Top or QCD**
- Move to **DNN-based flavor tagging** plays significant role in all sensitivity improvements highlighted today

# Flavor tagging: constituent-based algorithms



- Run 3 FTAG in ATLAS based on **transformer encoder** with **tracks as inputs**
  - **Attention mechanism** allows for incorporating **correlations between features from different tracks** (kinematics, IPs, lower-level tracking information)
- Jet flavor classification task aided by additional **physics-inspired auxiliary training objectives**
- **Shared structure** for GN1/GN2 [PHYS-PUB-2022-027] in resolved and GN2X [PHYS-PUB-2023-021] in boosted

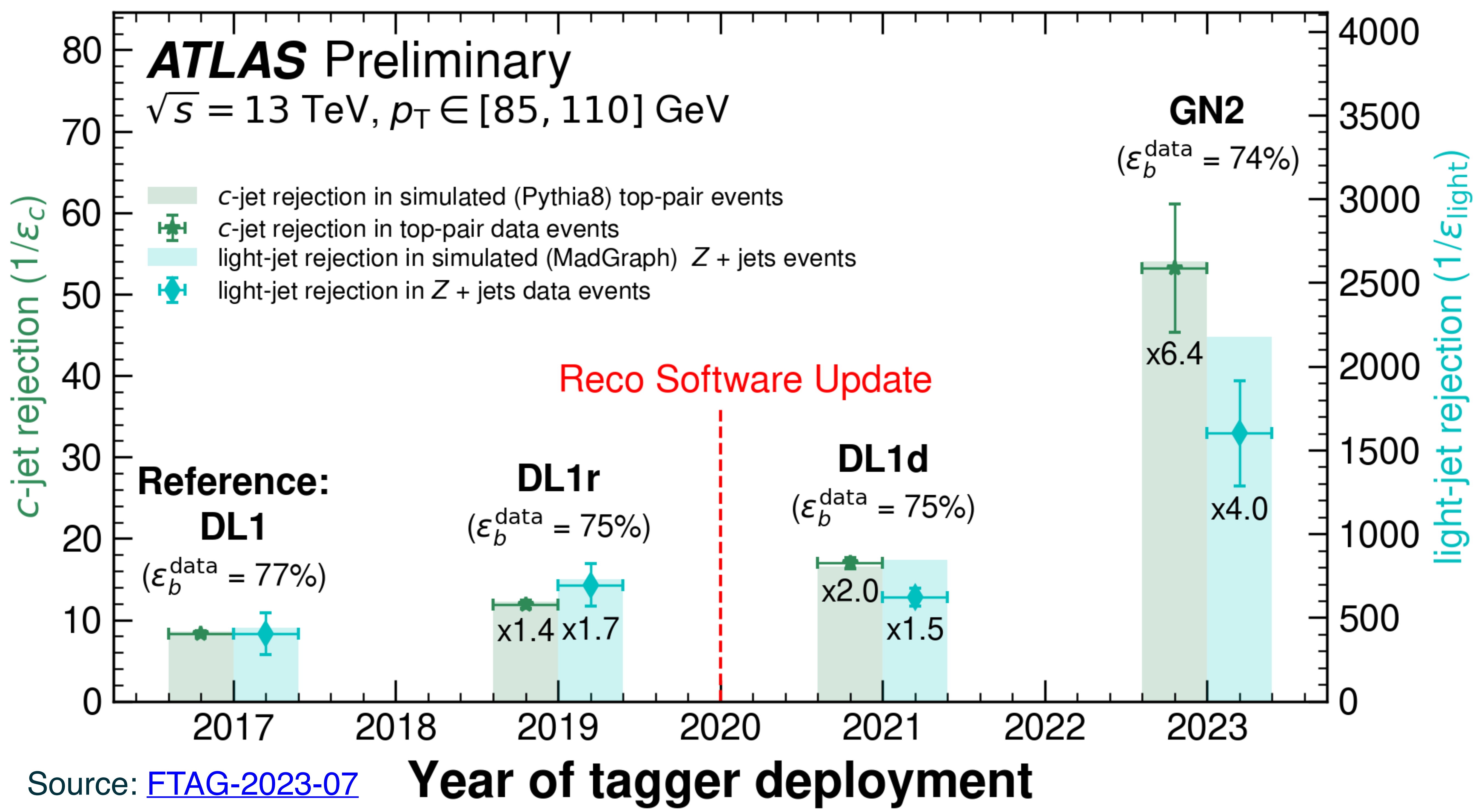



Source: ATL-PHYS-PUB-2022-027

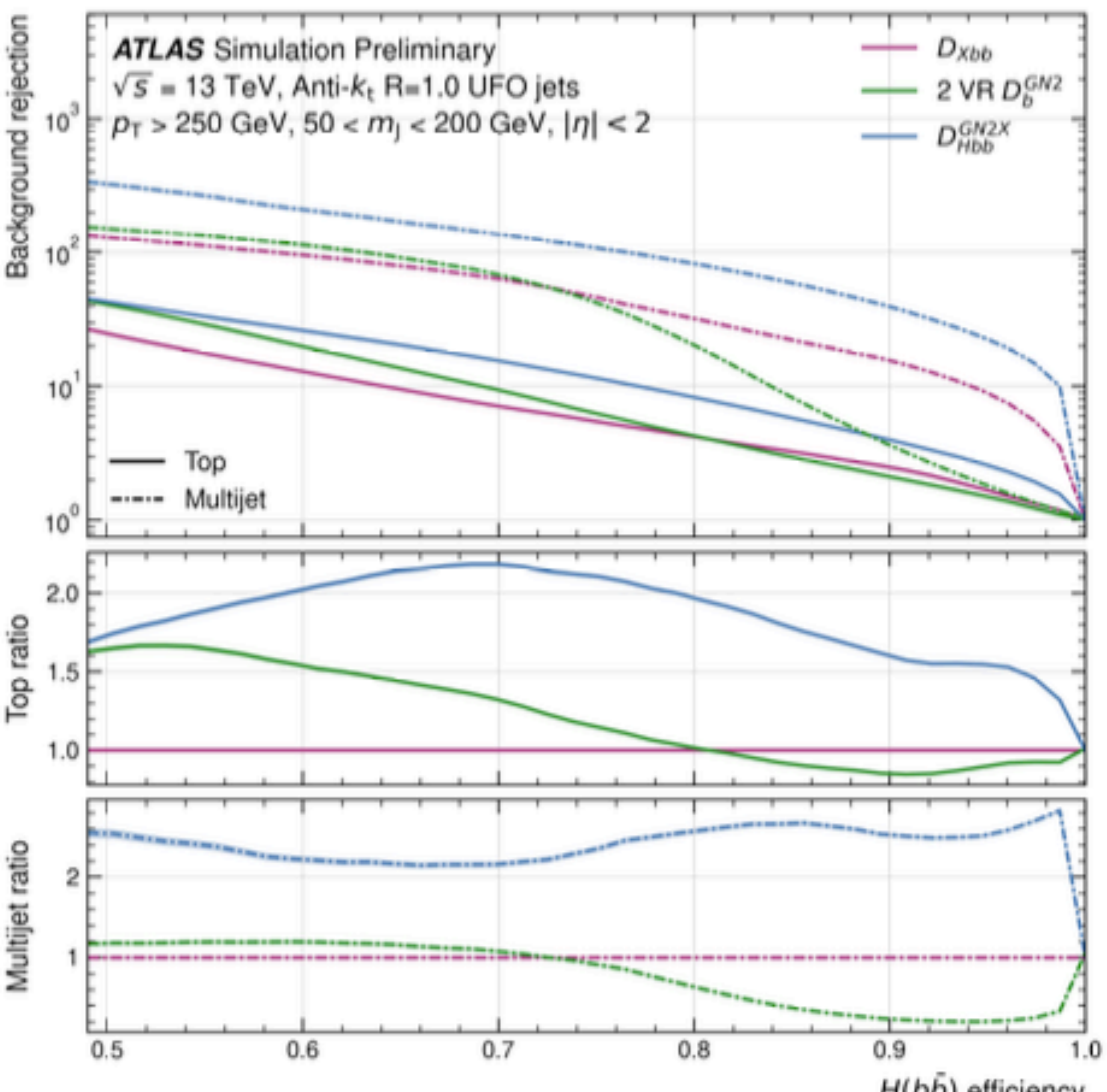
# Flavor tagging: performance comparisons



- Up to **4-times** increased background rejection for **bottom tagging** with GN2 compared to DL1r and **2/3-times for H(bb) tagging** with GN2X
- Similar improvements seen in **charm tagging**
  - GN2X is first algorithm in ATLAS to do **explicit H(cc) tagging**




Source: FTAG-2023-07

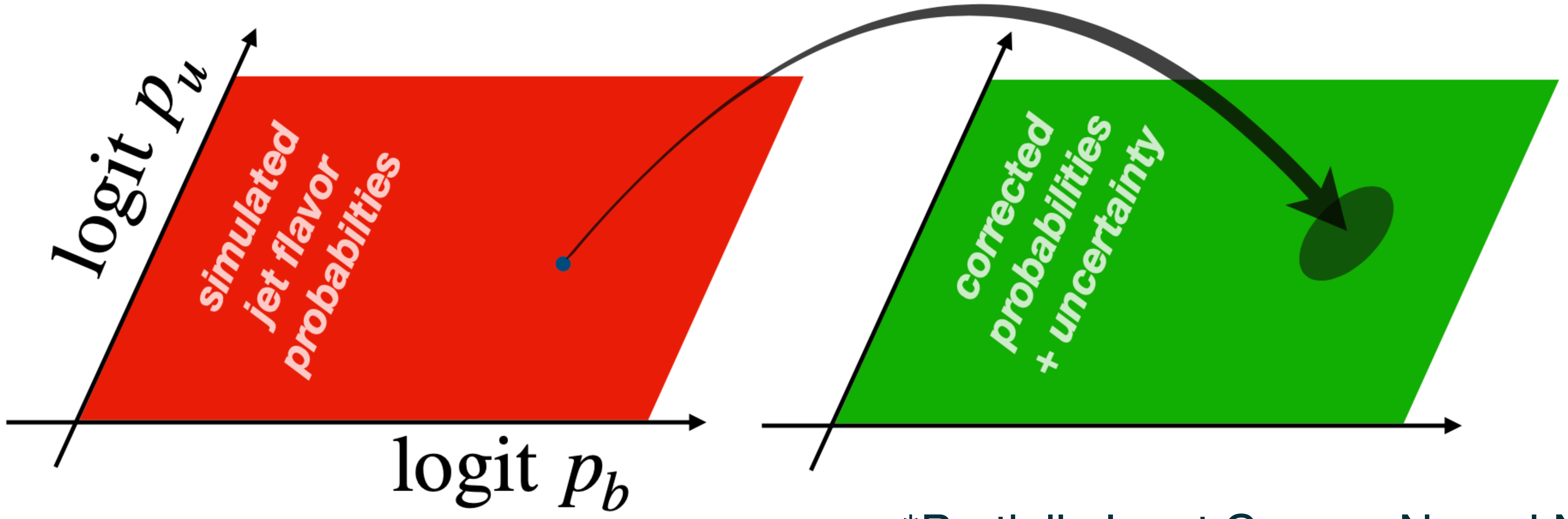


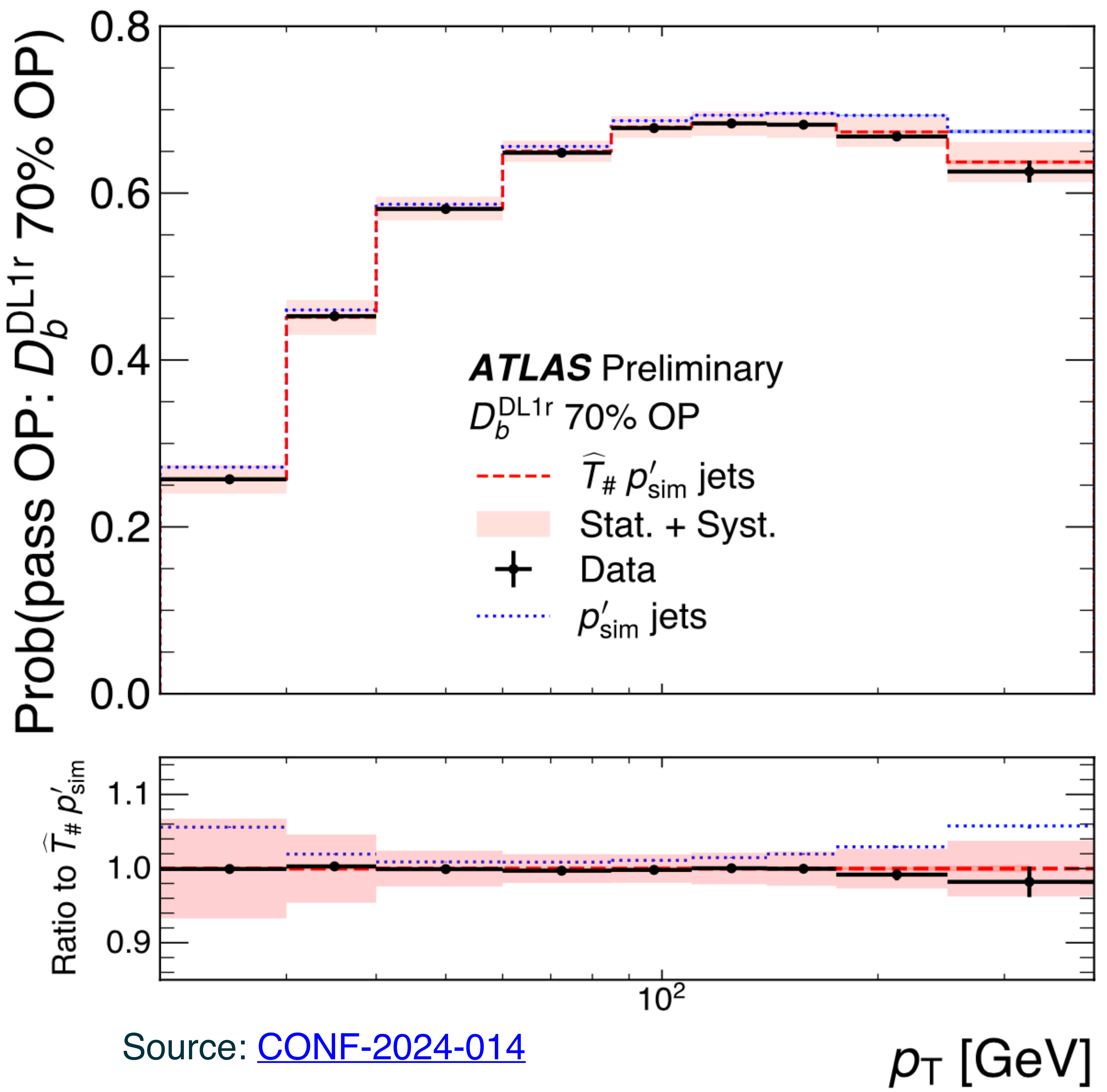

Source: ATL-PHYS-PUB-2023-021

# Flavor tagging: calibration via optimal transport



- FTAG algorithms **require individual calibrations** based on choice of discriminant
  - New method uses **optimal transport maps** to calibrate ($p_b$, $p_c$, $p_u$) continuously
- Idea is to derive **$p_T$ dependent mapping** $T_{p_T}$ which **matches data performance** and minimally alters the simulation (via euclidian distance metric)
  - Mapping is gradient of scalar complex function → use PICNN* to approximate
- **Method tested** [CONF-2024-014] on b-jets in leptonic $t\bar{t}$ decays with **good closure** across the board




Source: CONF-2024-014

*Partially Input Convex Neural Network

# Event selection: use of multi-variate discriminants

- Fitting BDT/DNN derived **multi-variate discriminants** gives increased S/B separation compared to single-variable based (i.e. $m_H$) fits
- MVA techniques bring **notable improvements** to $V(\to l\bar{l})\ H(\to b\bar{b}/c\bar{c})$ [HIGG-2020-20] and boosted VBF $HH(\to b\bar{b}b\bar{b})$ [HDBS-2022-02] analyses

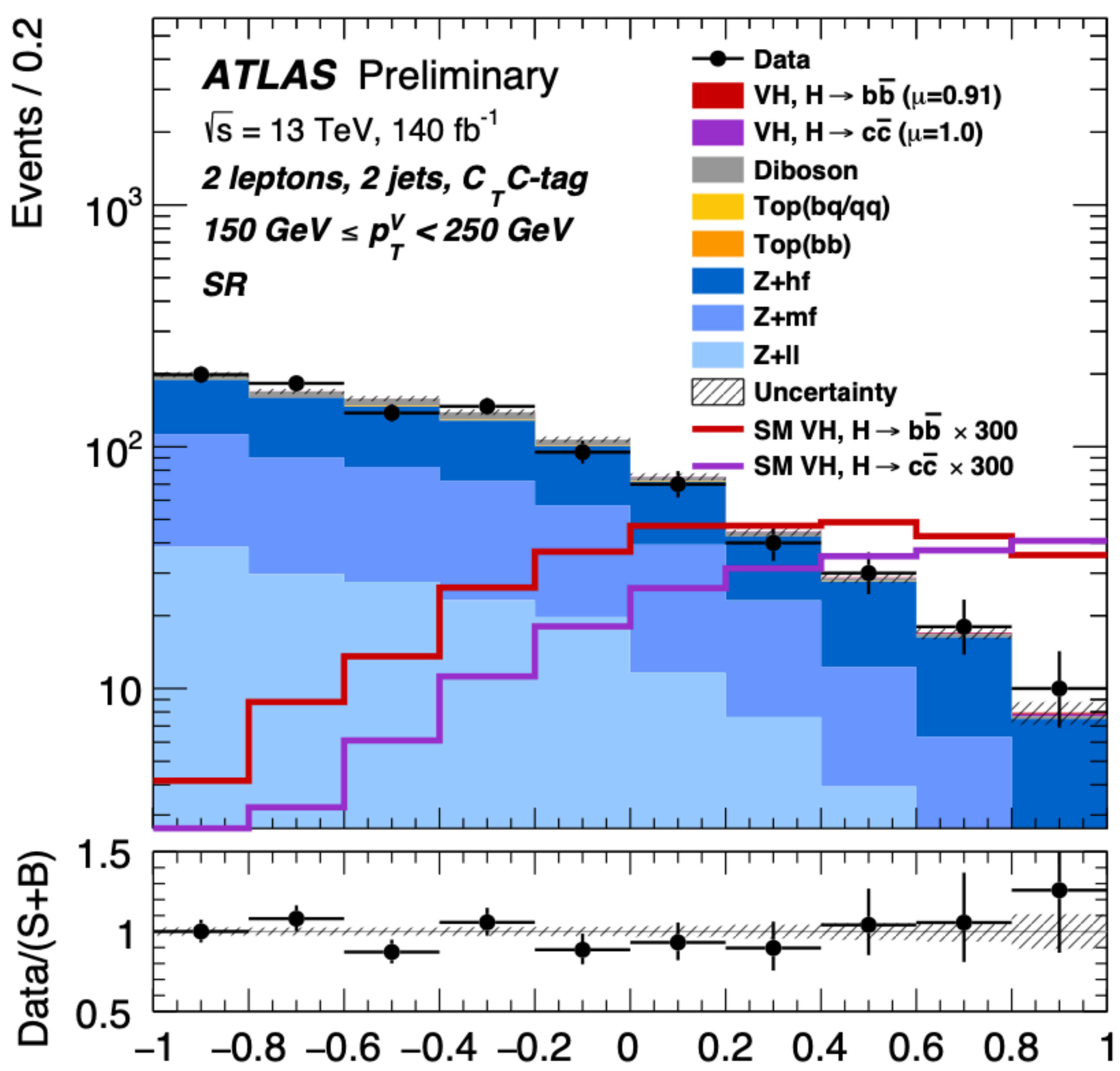
## $V(\to l\bar{l})\ H(\to b\bar{b}/c\bar{c})$

- **Most sensitive** $\mu_{VH(c\bar{c})}$ measurement
  - 11.5 (10.6) observed (expected) upper limit vs 26 (31)
- MVA responsible for **~50% of improvement**, FTAG for ~40%

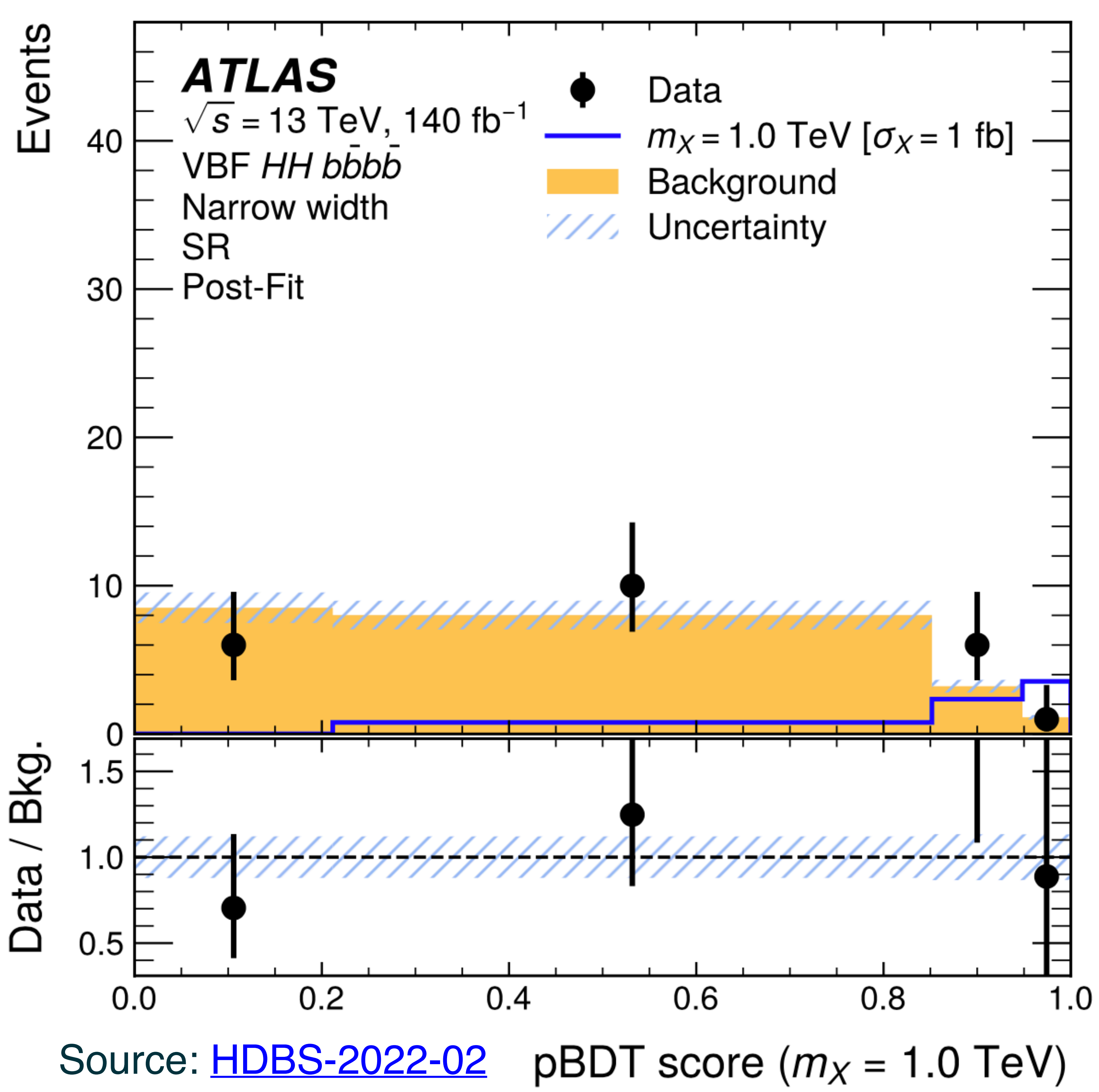
## $HH(\to b\bar{b}b\bar{b})$

- Resonant analysis uses **mass-parameterized BDTs** to test multiple mass hypotheses at once

See talk by *Marion Missio*




Source: HIGG-2020-20



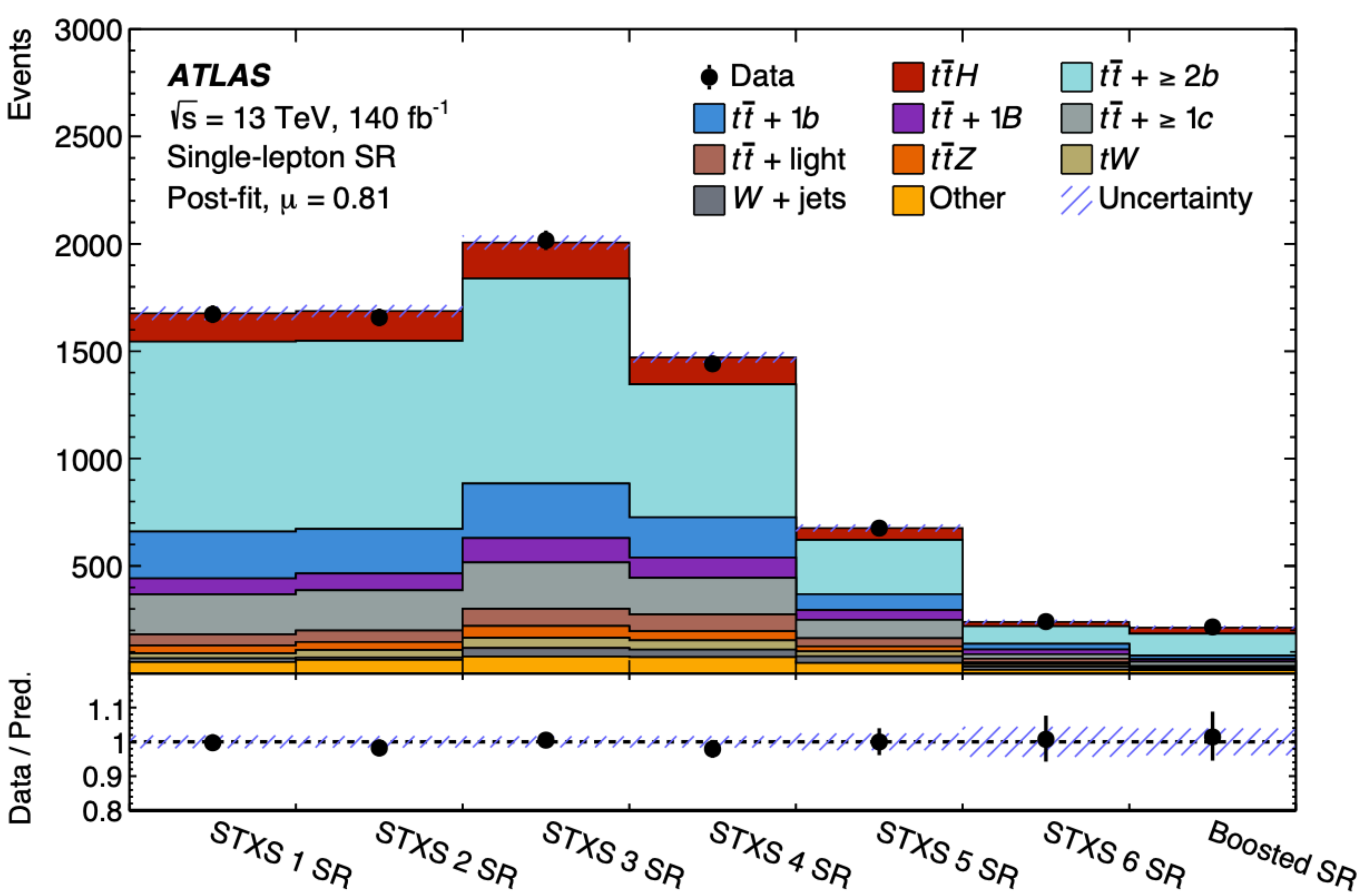

Source: HDBS-2022-02

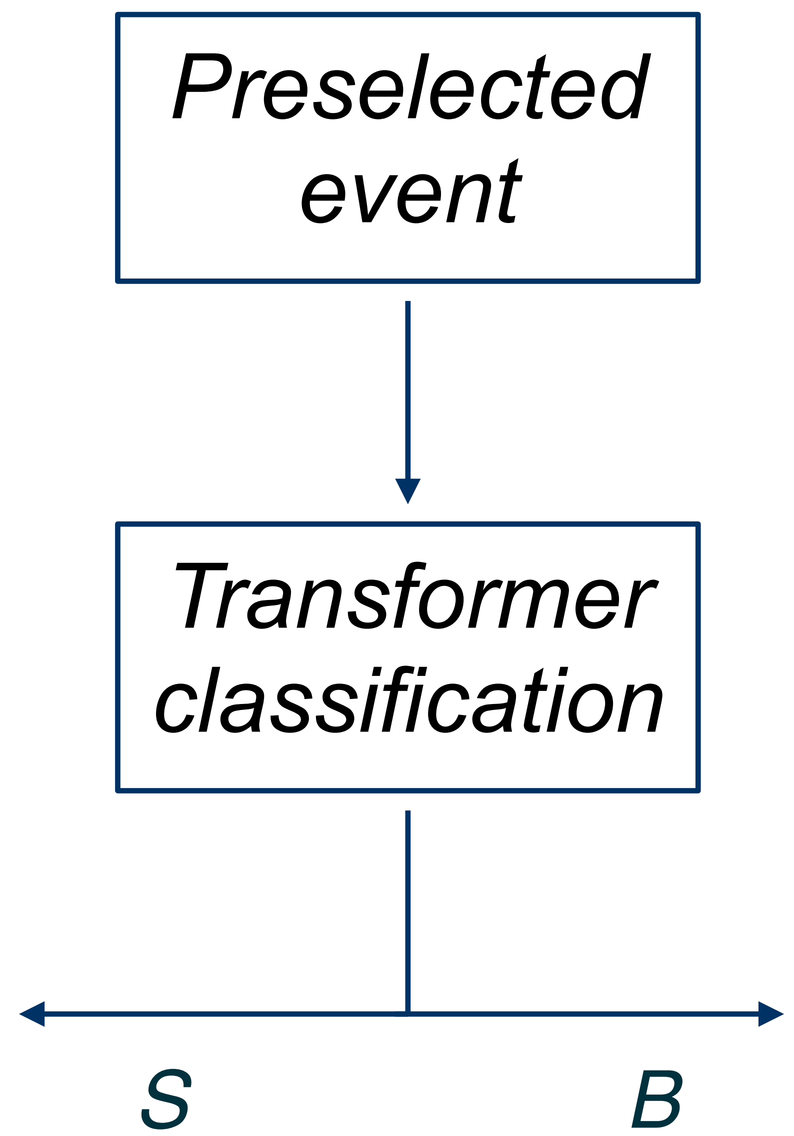
# Event selection: transformer-based categorization



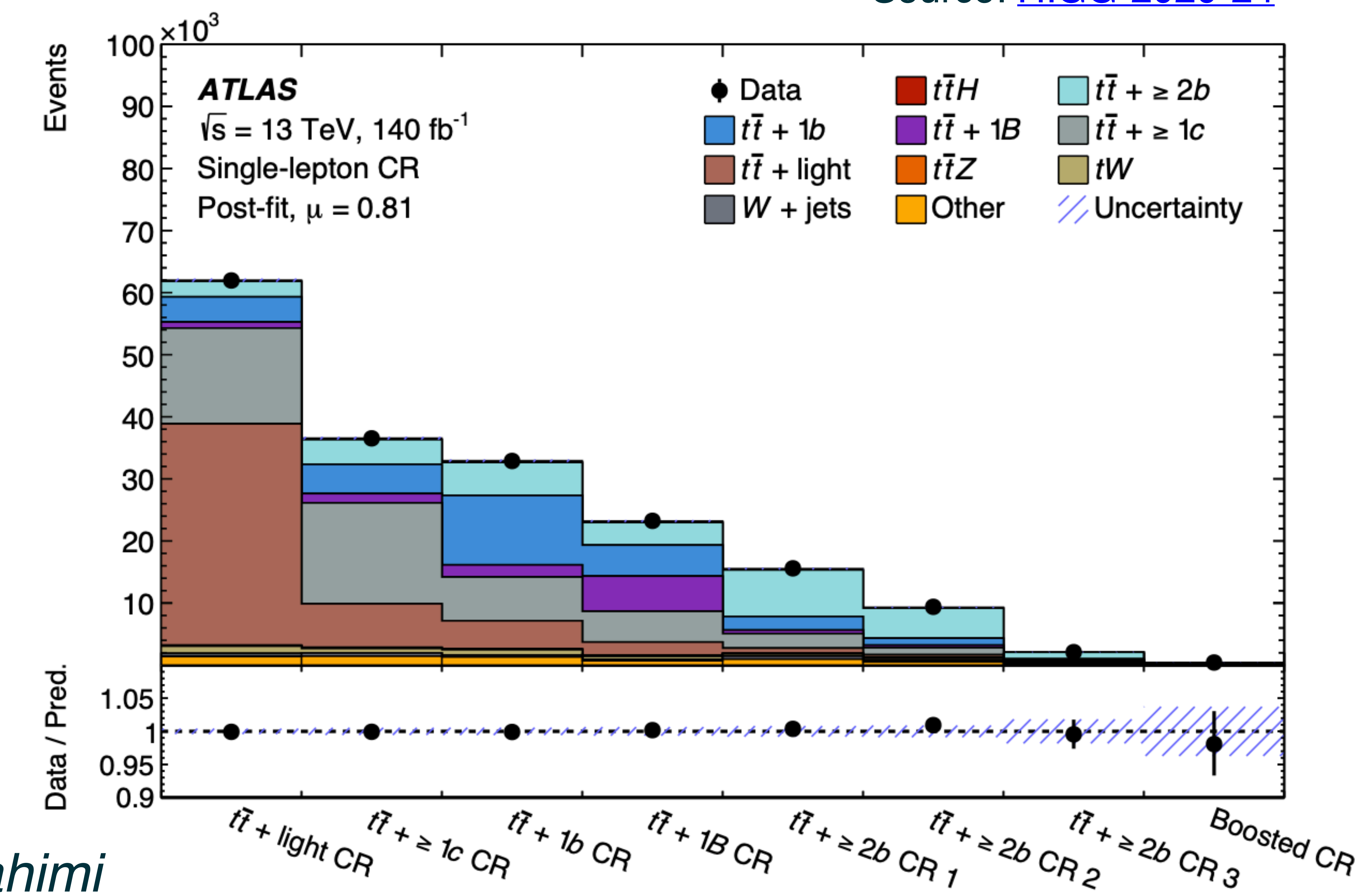
- New $t\bar{t}H(\to b\bar{b})$ result [HIGG-2020-24] categorizes events with two **Transformer-based** neural networks from **jet, lepton and MET input** information
  - One splits events between **signal and 5 $t\bar{t}$+jets background** categories, other identifies **Higgs candidate $p_T$** in signal events
- Main factor in improved $t\bar{t}H$ significance of 4.6 (5.4) observed (expected) vs 1.0 (2.7) in previous analysis with 94-97% $H(\to b\bar{b})$ purity among $t\bar{t}H$ events in SRs

Source: HIGG-2020-24




See talk by *Nihal Brahimi*



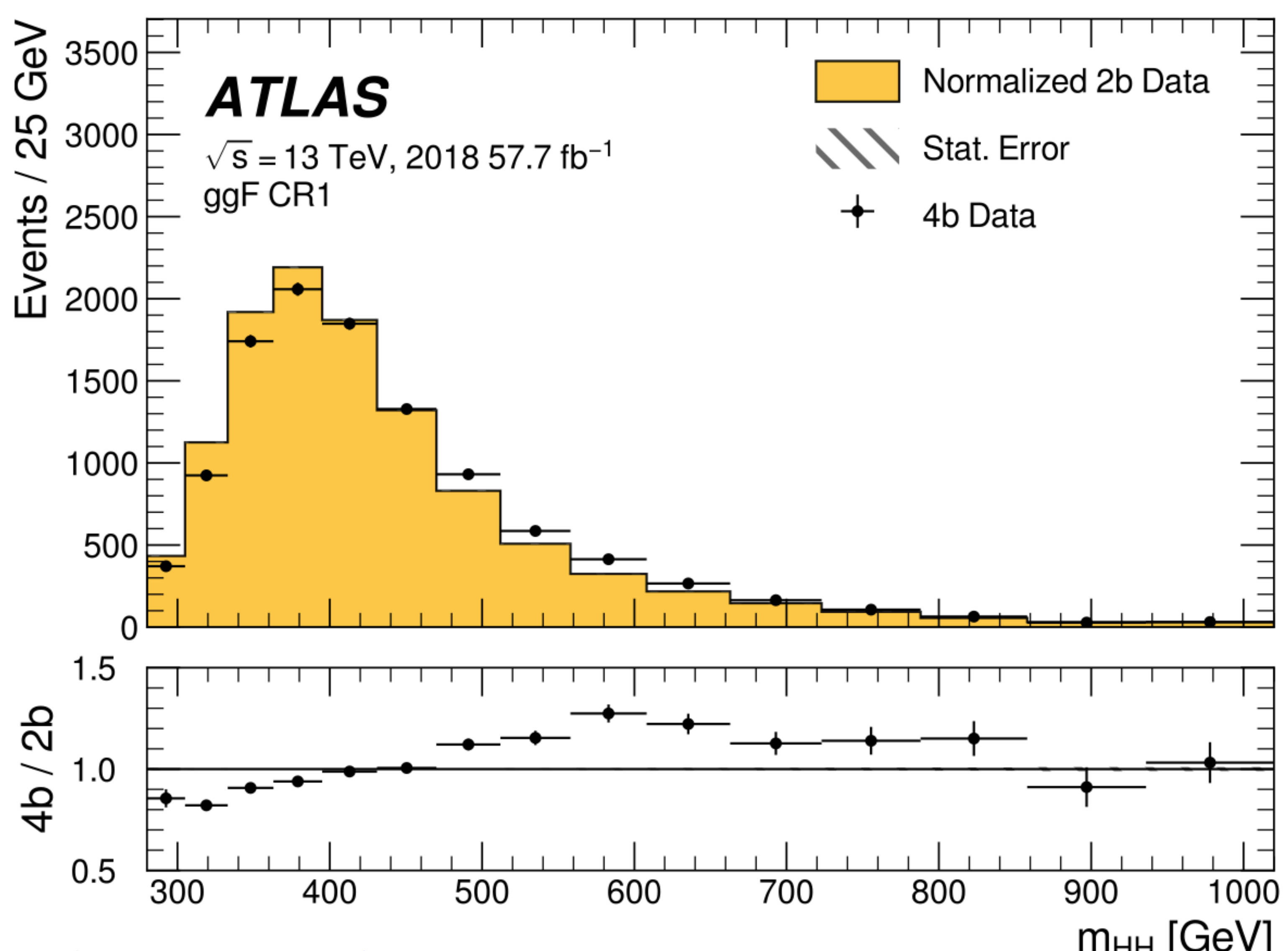
# Background modeling: density ratio estimation

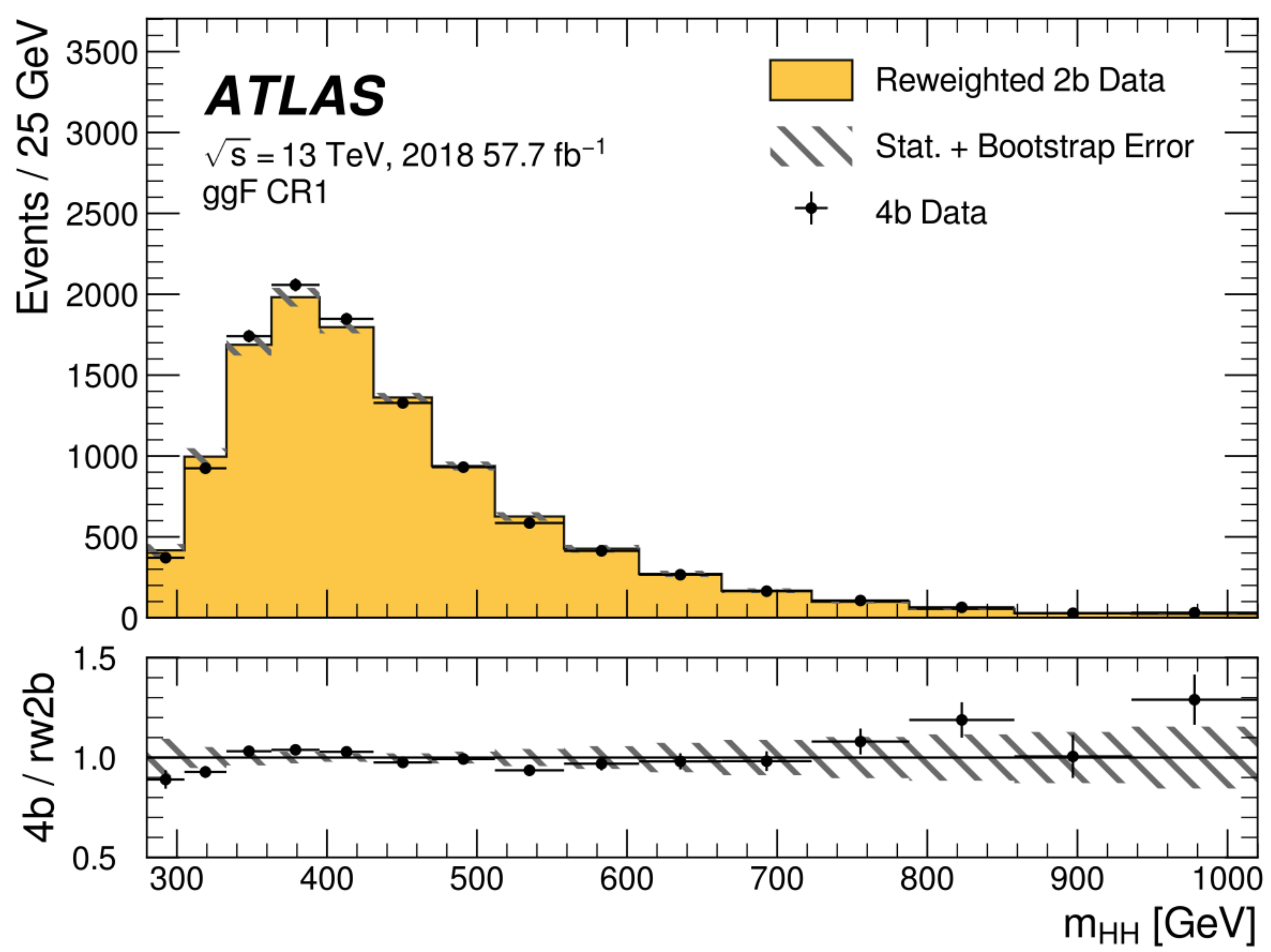


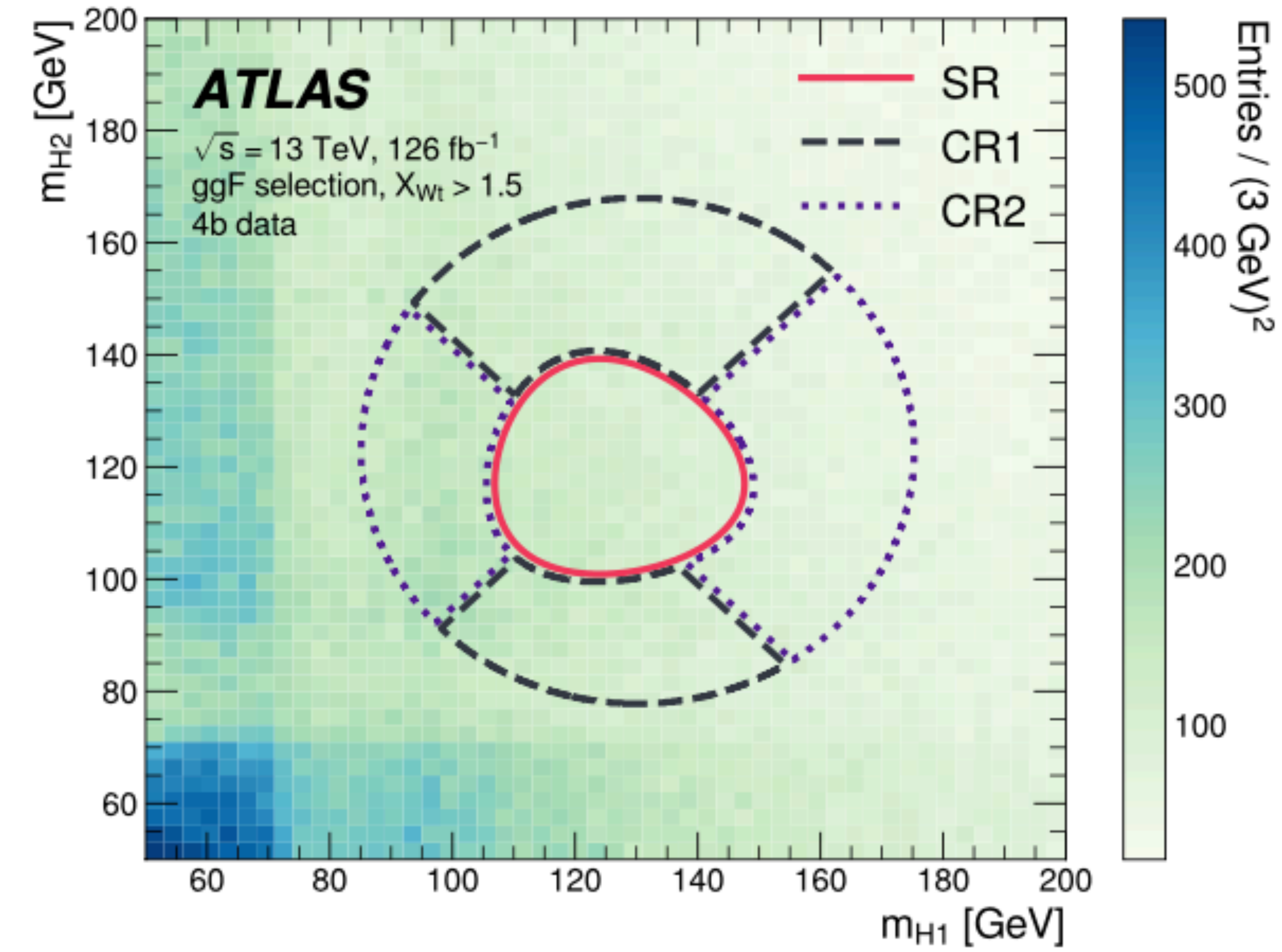
- ggF and VBF $HH(\rightarrow b\bar{b}b\bar{b})$ result [HDBS-2019-29] uses DNN for **density ratio estimation** to estimate dominant (~90%) **multi-jet background** based on data

  *reweighting function*

  - Background in SR modeled based on **2 b-tag events** with kinematic differences accounted for by reweighting function (2b → 4b) **learned in CRs**
- **Improved modeling of correlations** in kinematic variables leads to **better $m_{HH}$ modeling**, contributing to 30% improvement over lumi scaling



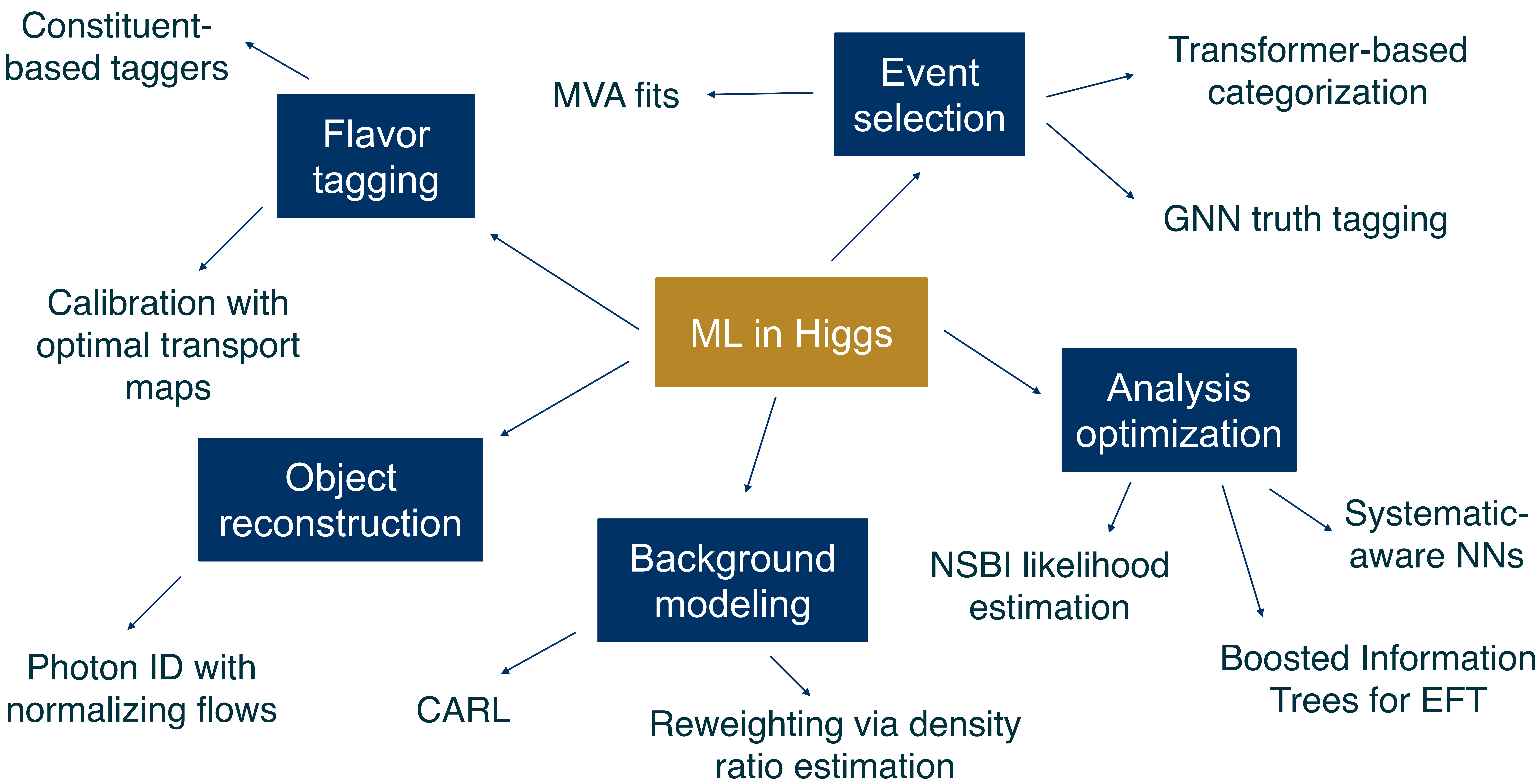

Source: HDBS-2019-29

# Brief overview of other techniques



- **MC-MC shape uncertainties** derived via Calibrated Likelihood Ratio Estimator (CARL) in $V(\to l\bar{l})\ H(\to b\bar{b}/c\bar{c})$ [CONF-2024-010]
  - DNN trained on **nominal and alternative MC** to distinguish the two
  - Output weights used to **reweight nominal** → better statistics than alternative
- **Neural simulation-based inference** (NSBI) used to estimate likelihood ratios from **unbinned events** by learning a **reweighting of one hypothesis** to another [CONF-2024-015]
  - See talk by *Jay Sandesara*
- **Weighted event selection** via GNN-based "truth tagging" in $V(\to l\bar{l})\ H(\to b\bar{b}/c\bar{c})$ [CONF-2024-010]
  - GNN trained on MC to predict probability that event **passes given FTAG selection**
  - All **MC events kept and weighted**, improving background modeling via increased statistics

# Today’s highlights



# What does the future hold?



- Machine learning is now a **mainstay of Higgs physics** in ATLAS and CMS
- Evolution characterized by **larger and more complex models** working with **lower-level inputs**
  - **Physics-inspired techniques** are gaining a lot of traction
  - Gradual **increase in training statistics** often still lead to performance gains
- Use of ML has **expanded greatly** beyond just object reconstruction and event selection
- Future holds many interesting possibilities as we move towards HL-LHC
  - **On the horizon**: global particle flow, physics foundation models, ML for unfolding, …

*Lots of exciting ML-related improvements to look forward to in run 3!*

# Thank you!

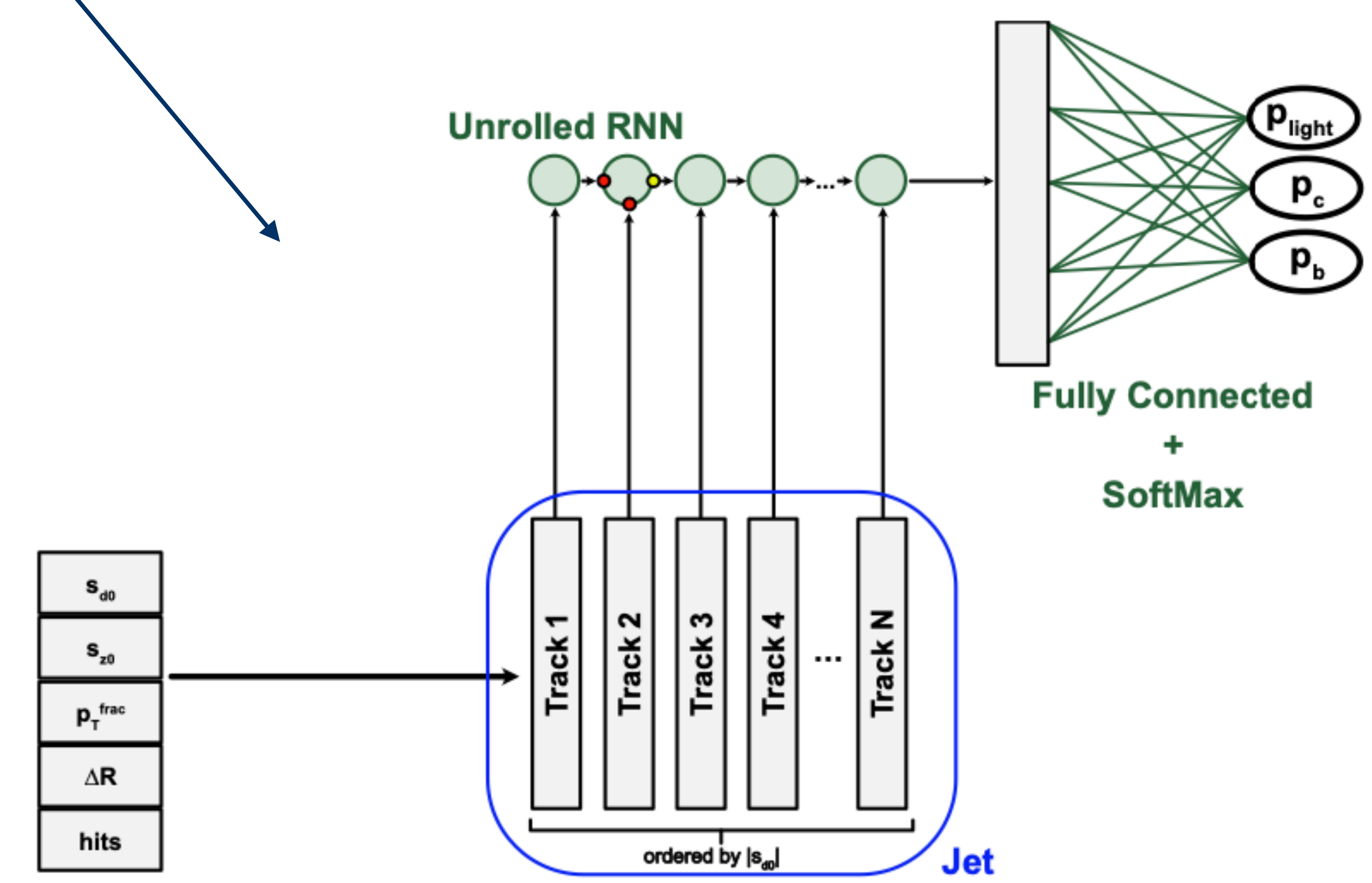
*This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 2146752. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.*

# Backup

# DL1r details



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

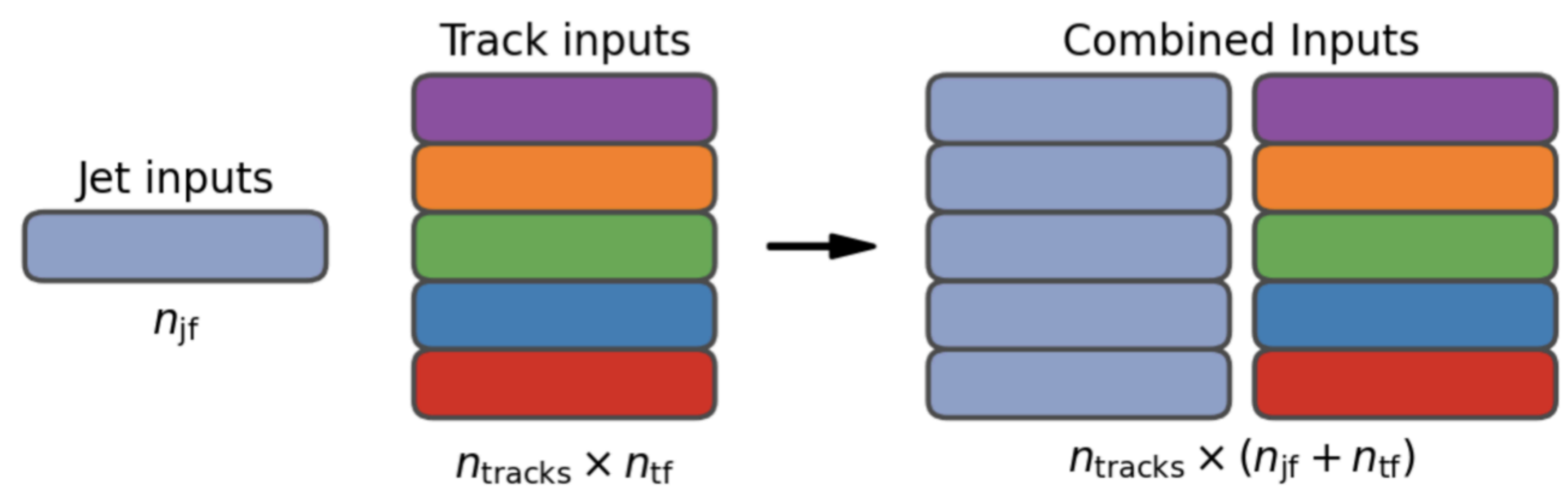



Source: FTAG-2019-07

# GN1/GN2 model details



| Jet Input | Description |
|---|---|
| $p_T$ | Jet transverse momentum |
| $\eta$ | Signed jet pseudorapidity |
| **Track Input** | **Description** |
| $q/p$ | Track charge divided by momentum (measure of curvature) |
| d$\eta$ | Pseudorapidity of the track, relative to the jet $\eta$ |
| d$\phi$ | Azimuthal angle of the track, relative to the jet $\phi$ |
| $d_0$ | Closest distance from the track to the PV in the longitudinal plane |
| $z_0 \sin\theta$ | Closest distance from the track to the PV in the transverse plane |
| $\sigma(q/p)$ | Uncertainty on $q/p$ |
| $\sigma(\theta)$ | Uncertainty on track polar angle $\theta$ |
| $\sigma(\phi)$ | Uncertainty on track azimuthal angle $\phi$ |
| $s(d_0)$ | Lifetime signed transverse IP significance |
| $s(z_0)$ | Lifetime signed longitudinal IP significance |
| nPixHits | Number of pixel hits |
| nSCTHits | Number of SCT hits |
| nIBLHits | Number of IBL hits |
| nBLHits | Number of B-layer hits |
| nIBLShared | Number of shared IBL hits |
| nIBLSplit | Number of split IBL hits |
| nPixShared | Number of shared pixel hits |
| nPixSplit | Number of split pixel hits |
| nSCTShared | Number of shared SCT hits |
| nPixHoles | Number of pixel holes |
| nSCTHoles | Number of SCT holes |
| leptonID | Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep) |




Source: ATL-PHYS-PUB-2022-027

| Truth Origin | Description |
|---|---|
| Pileup | From a $pp$ collision other than the primary interaction |
| Fake | Created from the hits of multiple particles |
| Primary | Does not originate from any secondary decay |
| fromB | From the decay of a $b$-hadron |
| fromBC | From a $c$-hadron decay, which itself is from the decay of a $b$-hadron |
| fromC | From the decay of a $c$-hadron |
| OtherSecondary | From other secondary interactions and decays |

Track origin classification labels

GN1/GN2 input features

GN1 vs GN2 differences

Source: FTAG-2023-01

| Type | Name | GN1 | GN2 |
|---|---|---|---|
| Hyperparameter | Trainable parameters | 0.8M | 1.5M |
| Hyperparameter | Learning rate | 1e−3 | OneCycle LRS (max LR 4e−5) |
| Hyperparameter | GNN Layers | 3 | 6 |
| Hyperparameter | Attention Heads | 2 | 8 |
| Hyperparameter | Embed. dim | 128 | 192 |
| Architectural | Attention type | GATv2 | ScaledDotProduct |
| Architectural | Dense update | No | Yes (dim 256) |
| Architectural | Separate value projection | No | Yes |
| Architectural | LayerNorm + Dropout | No | Yes |
| Inputs | Num. training jets | 30M | 192M |

# GN2 b-tagging performance

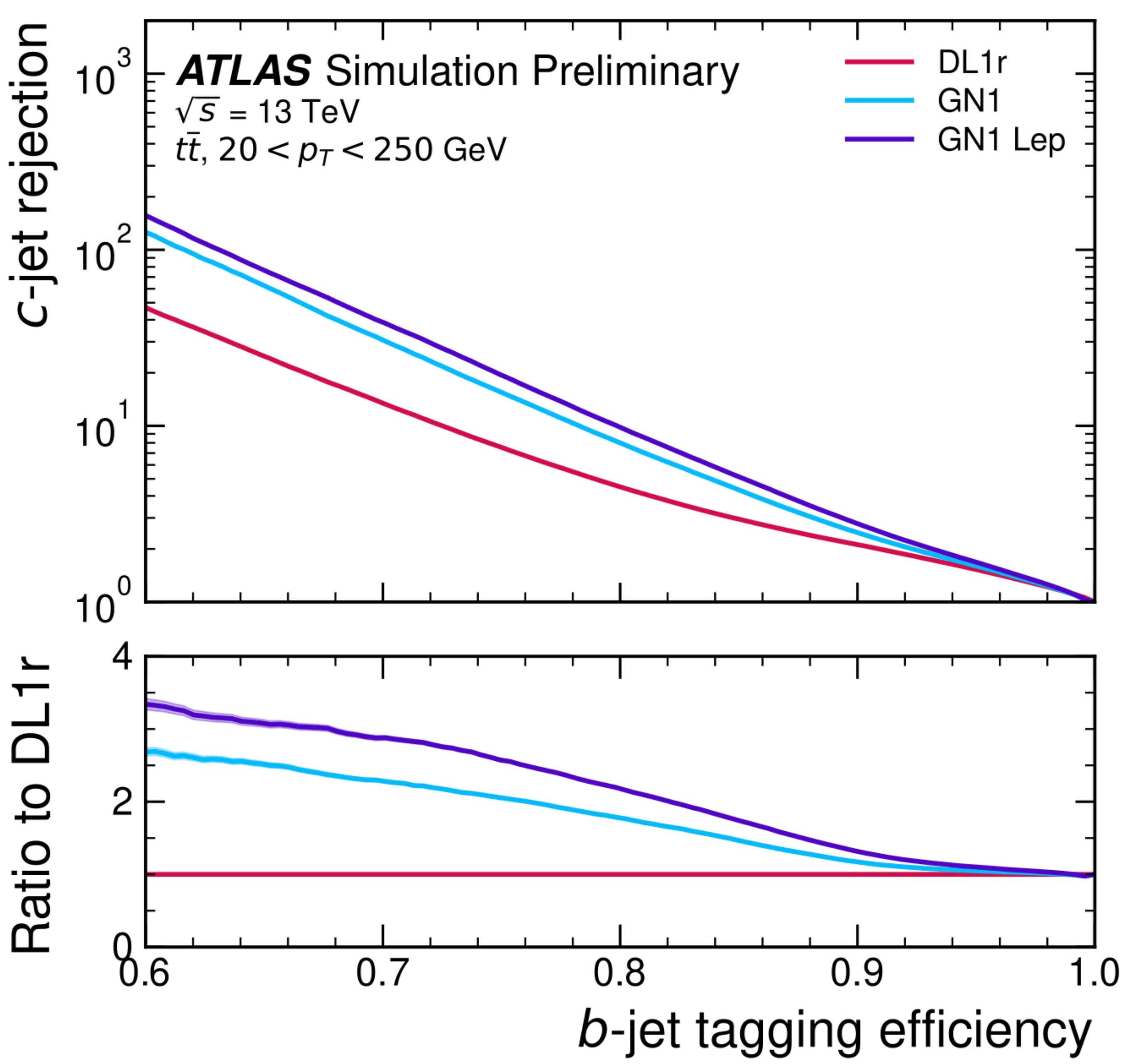


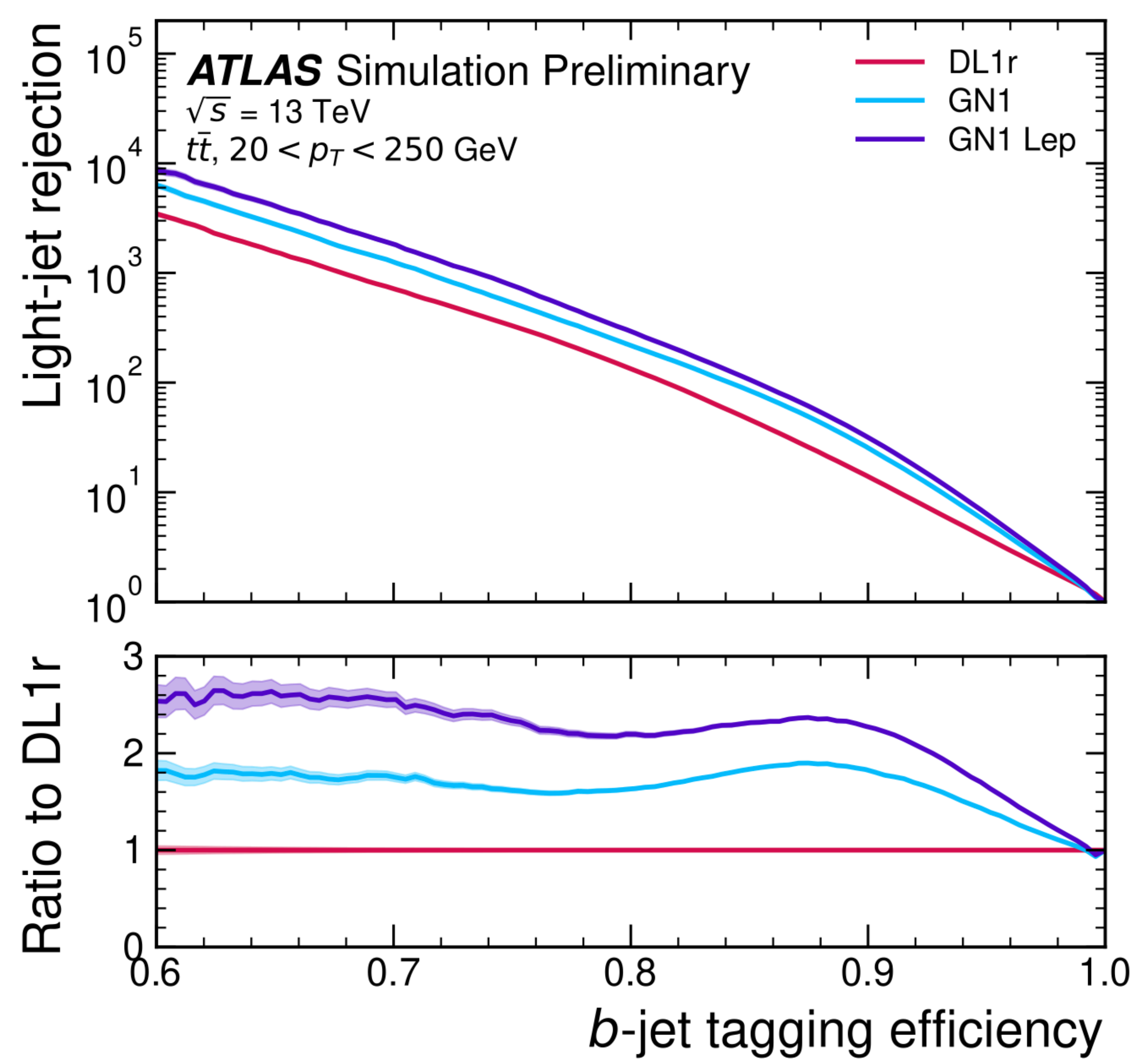
Source: [FTAG-2023-01]

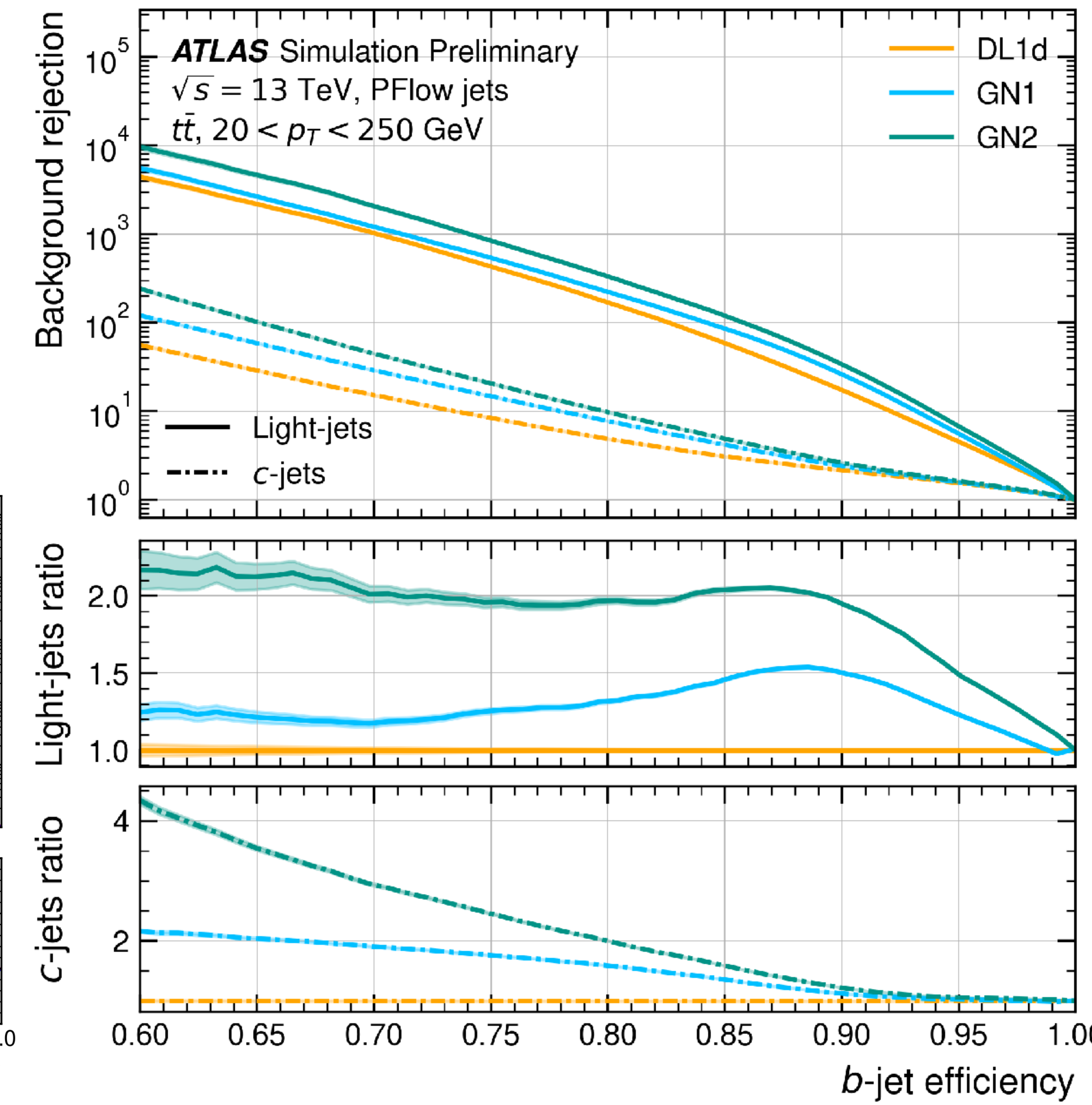
GN2 vs. GN1

GN1 vs. DL1r

Source: [ATL-PHYS-PUB-2022-027]



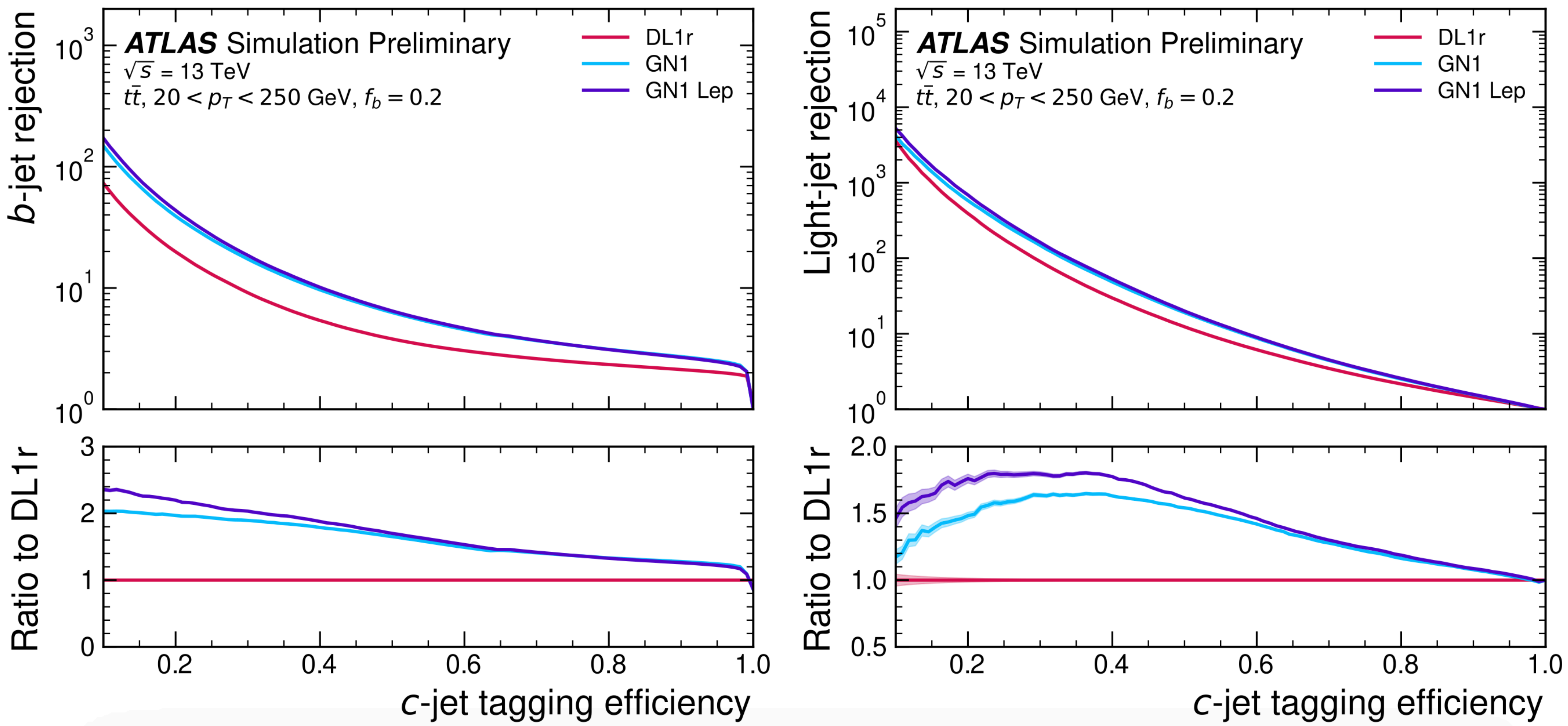
# GN2 c-tagging performance



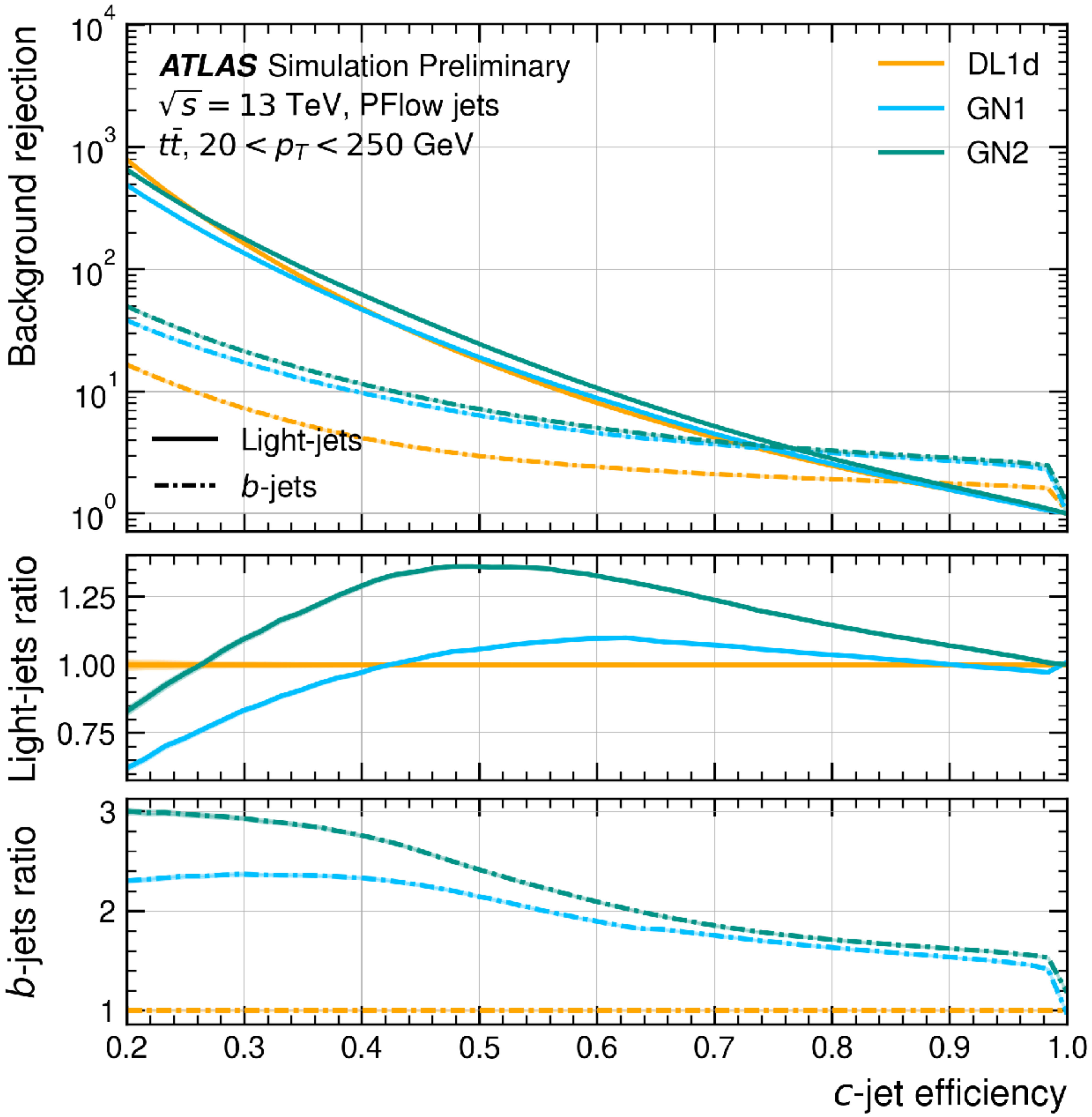
GN2 vs. GN1

GN1 vs. DL1r

Source: ATL-PHYS-PUB-2022-027




Source: FTAG-2023-01



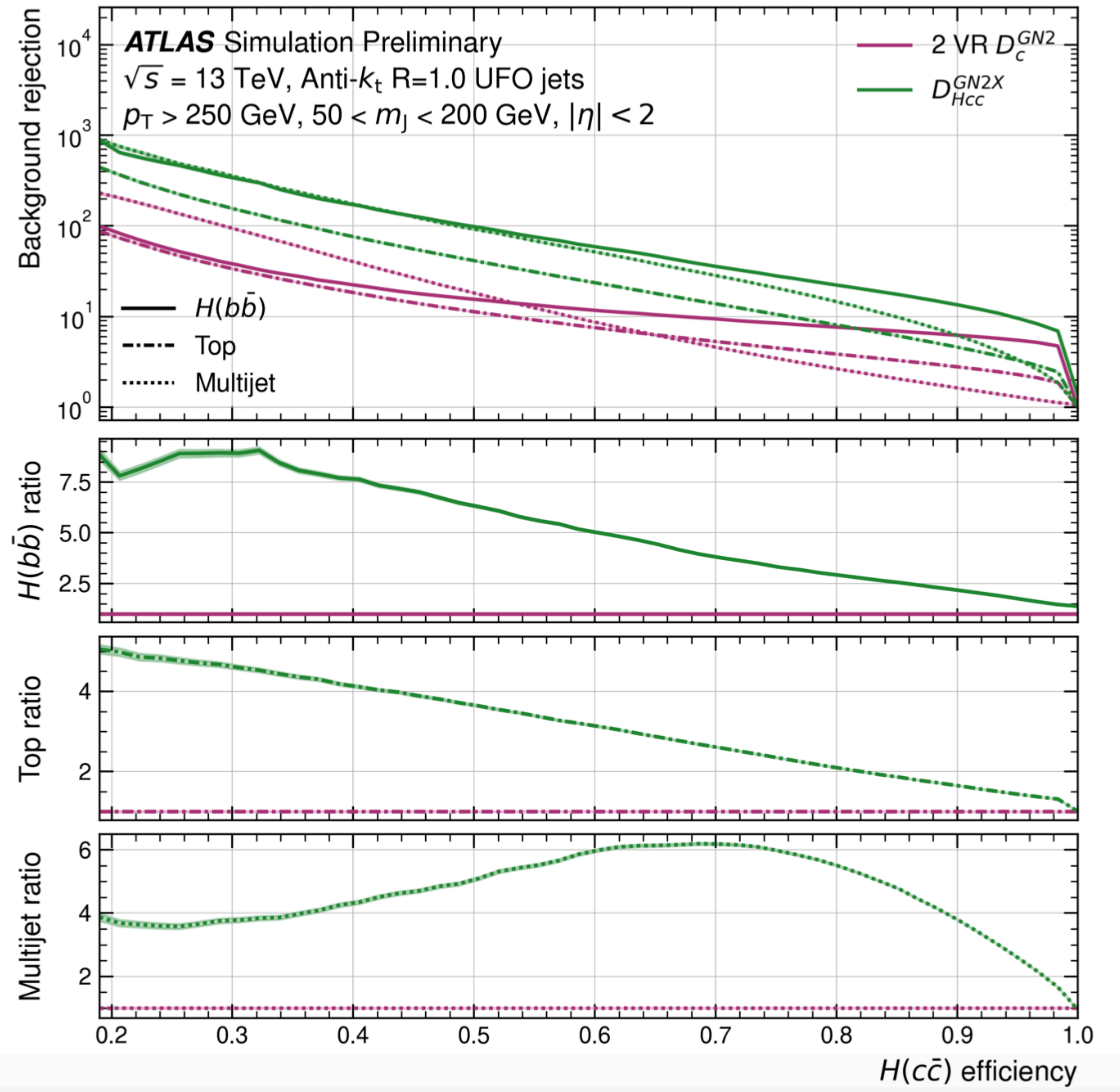
# GN2X inputs and H(cc)-tagging performance



| Jet Input | Description |
|---|---|
| $p_{\mathrm{T}}$ | Large-$R$ jet transverse momentum |
| $\eta$ | Signed large-$R$ jet pseudorapidity |
| mass | Large-$R$ jet mass |
| **Track Input** | **Description** |
| $q/p$ | Track charge divided by momentum (measure of curvature) |
| $\mathrm{d}\eta$ | Pseudorapidity of track relative to the large-$R$ jet $\eta$ |
| $\mathrm{d}\phi$ | Azimuthal angle of the track, relative to the large-$R$ jet $\phi$ |
| $d_0$ | Closest distance from track to primary vertex (PV) in the transverse plane |
| $z_0 \sin\theta$ | Closest distance from track to PV in the longitudinal plane |
| $\sigma(q/p)$ | Uncertainty on $q/p$ |
| $\sigma(\theta)$ | Uncertainty on track polar angle $\theta$ |
| $\sigma(\phi)$ | Uncertainty on track azimuthal angle $\phi$ |
| $s(d_0)$ | Lifetime signed transverse IP significance |
| $s(z_0 \sin\theta)$ | Lifetime signed longitudinal IP significance |
| nPixHits | Number of pixel hits |
| nSCTHits | Number of SCT hits |
| nIBLHits | Number of IBL hits |
| nBLHits | Number of B-layer hits |
| nIBLShared | Number of shared IBL hits |
| nIBLSplit | Number of split IBL hits |
| nPixShared | Number of shared pixel hits |
| nPixSplit | Number of split pixel hits |
| nSCTShared | Number of shared SCT hits |

Source: ATL-PHYS-PUB-2023-021

GN2X input features



# Optimal transport calibration method

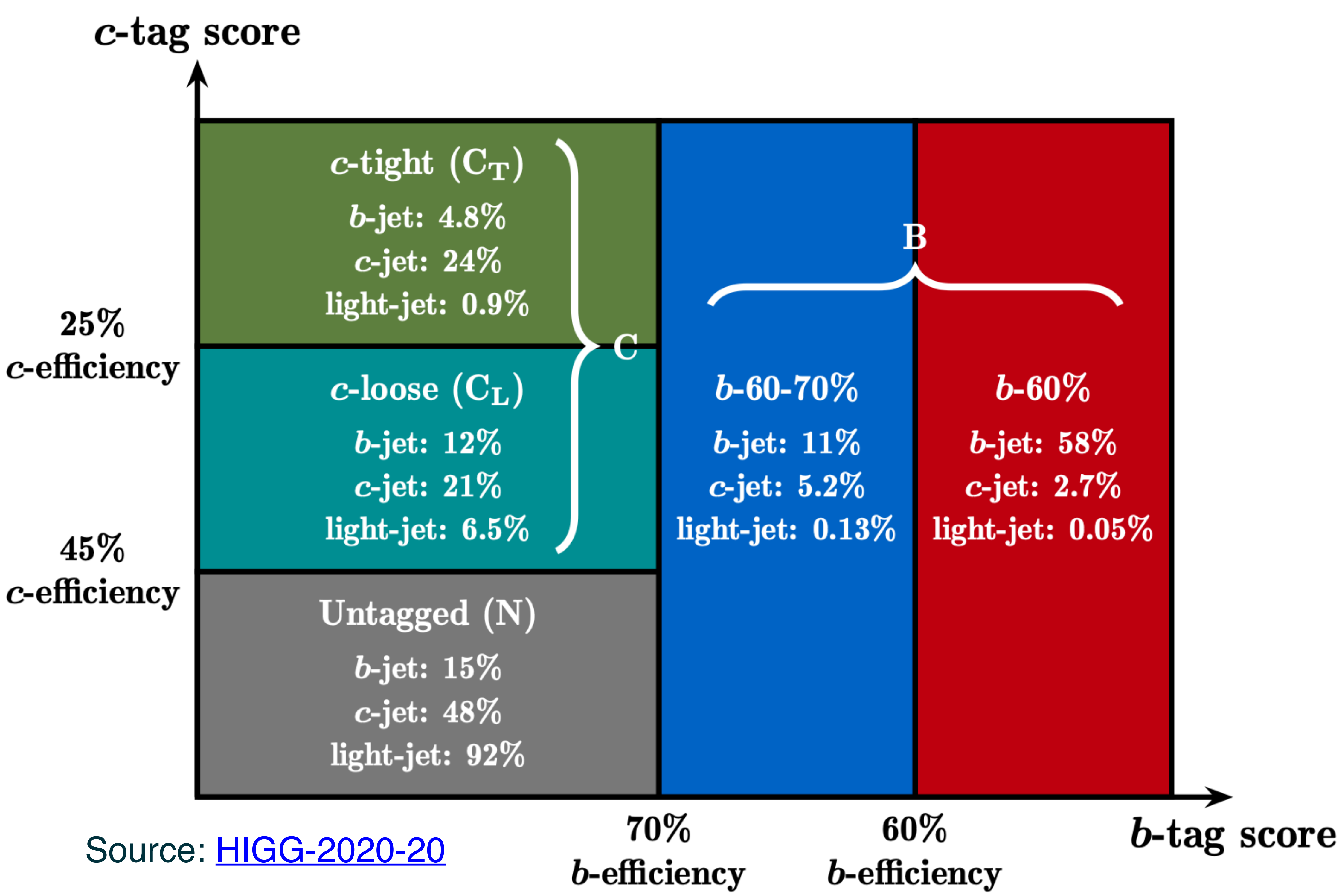


- Optimal transport maps **derived for** $q_i = \mathrm{logit}\ p_i$ where $\mathrm{logit}\ p_i = \log \dfrac{p_i}{1-p_i}$
  - Defined by $\hat{T}_{p_\mathrm{T}} = \arg\inf_{T_{p_\mathrm{T}}} \int_{\vec{q}\in\vec{Q}} c(\vec{q}, T_{p_\mathrm{T}}\,\vec{q}) p_\mathrm{sim}(\vec{q}|p_\mathrm{T}) d\vec{q};\ \mathrm{s.t.}\ (T_{p_\mathrm{T}})_\#\, p_\mathrm{sim}(\vec{q}|p_\mathrm{T}) \approx p_\mathrm{data}(\vec{q}|p_\mathrm{T})$ with distance metric $c^2(\vec{x},\vec{y}) = (\vec{x}-\vec{y})^2$ (unique solution)
- **Partially Input Convex Neural Networks** (PICNNs) used to approximate optimal mapping with loss (maximize f, minimize g)

$$\mathcal{L}(f,g) = \frac{1}{N_\mathrm{data}} \sum_{\vec{q},p_\mathrm{T}\sim p_\mathrm{data}} f(\vec{q},p_\mathrm{T}) + \frac{1}{N_\mathrm{sim}} \sum_{\vec{q},p_\mathrm{T}\sim p'_\mathrm{sim}} \vec{q}\cdot\vec{\nabla} g(\vec{q},p_\mathrm{T}) - f(\vec{\nabla} g(\vec{q},p_\mathrm{T}),p_\mathrm{T})$$

  - Input data for simulation sampled from **pT corrected distribution** (equal in data and MC) $p'_\mathrm{sim}(\vec{q},\ p_\mathrm{T}) \equiv (f_\mathrm{sig}(p_\mathrm{T}) p_\mathrm{sig}(\vec{q}|p_\mathrm{T}) + (1-f_\mathrm{sig}(p_\mathrm{T})) p_\mathrm{bkg}(\vec{q}|p_\mathrm{T}))\ p_\mathrm{data}(p_\mathrm{T})$
  - Requires understanding of **b-jet fractions** (i.e. $f_\mathrm{sig}(p_\mathrm{T})$ derived via neural density estimation) and **flavor probability distribution components** (i.e. $p_\mathrm{sig}(\vec{q}|p_\mathrm{T})$ derived via conditional normalizing flows)

- **3 sets of BDTs**: $\mathrm{BTD_{VH}}$, $\mathrm{BDT_{VZ}}$, $\mathrm{BDT_{CRLow}}$
  - BDT for VZ used in cross check analysis measuring VZ(bb/cc)
  - CRLow BDT applied in low-$\Delta R$ CR to separate V+jets from Top background



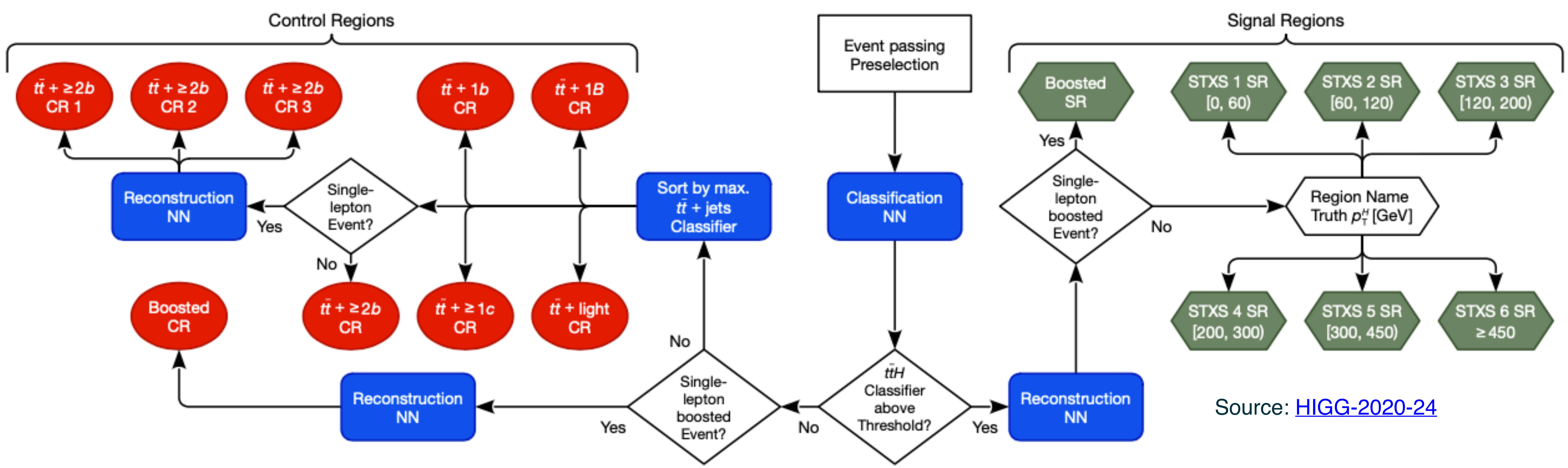

Source: HIGG-2020-20

| | Resolved $VH, H \to b\bar{b}, c\bar{c}$ | | | Boosted $VH, H \to b\bar{b}$ | | |
|---|---|---|---|---|---|---|
| Variable | 0-lepton | 1-lepton | 2-lepton | 0-lepton | 1-lepton | 2-lepton |
| $m_H$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $m_{j_1 j_2 j_3}$ | ✓ | ✓ | ✓ | | | |
| $p_T^{j_1}$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $p_T^{j_2}$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $p_T^{J_3}$ | | | | ✓ | ✓ | ✓ |
| $\sum p_T^{j_i}, i > 2$ | ✓ | ✓ | ✓ | | | |
| $\mathrm{bin}_{D_{\mathrm{DL1r}}}(j_1)$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\mathrm{bin}_{D_{\mathrm{DL1r}}}(j_2)$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $p_T^V$ | $\equiv E_T^{\mathrm{miss}}$ | ✓ | ✓ | $\equiv E_T^{\mathrm{miss}}$ | ✓ | ✓ |
| $E_T^{\mathrm{miss}}$ | ✓ | ✓ | | ✓ | ✓ | |
| $E_T^{\mathrm{miss}}/\sqrt{S_T}$ | | | ✓ | | | |
| $\lvert\Delta\phi(\boldsymbol{V},\boldsymbol{H})\rvert$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\lvert\Delta y(\boldsymbol{V},\boldsymbol{H})\rvert$ | | ✓ | ✓ | | ✓ | ✓ |
| $\Delta R(\boldsymbol{j_1},\boldsymbol{j_2})$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\min[\Delta R(\boldsymbol{j_i}, \boldsymbol{j_1}\ \mathrm{or}\ \boldsymbol{j_2})], i > 2$ | ✓ | ✓ | | | | |
| $N$(track-jets in $J$) | | | | ✓ | ✓ | ✓ |
| $N$(add. small-$R$ jets) | | | | ✓ | ✓ | ✓ |
| colour ring | | | | ✓ | ✓ | ✓ |
| $\lvert\Delta\eta(\boldsymbol{j_1},\boldsymbol{j_2})\rvert$ | ✓ | | | | | |
| $H_T + E_T^{\mathrm{miss}}$ | ✓ | | | | | |
| $m_T^W$ | | ✓ | | | | |
| $m_{\mathrm{top}}$ | | ✓ | | | | |
| $\min[\Delta\phi(\boldsymbol{\ell}, \boldsymbol{j_1}\ \mathrm{or}\ \boldsymbol{j_2})]$ | | ✓ | | | | |
| $p_T^{\ell}$ | | | | | ✓ | |
| $(p_T^{\ell} - E_T^{\mathrm{miss}})/p_T^V$ | | | | | ✓ | |
| $m_{\ell\ell}$ | | | ✓ | | | |
| $\cos\theta^*(\boldsymbol{\ell^-},\boldsymbol{V})$ | | | ✓ | | | ✓ |

# $t\bar{t}H(\to b\bar{b})$ transformer model



- $t\bar{t} + jets$ background is **split into 5 categories** based on **jets not from $t\bar{t}$ decay**: $t\bar{t} + \geq 2b$ (2 or more b-jets), $t\bar{t} + 1B$ (1 b-jet with 2 matched hadrons), $t\bar{t} + 1b$ (1 b-jet with 1 matched hadron), $t\bar{t} + \geq 1c$ (no b-jets, but at least 1 c-jet) and $t\bar{t} + light$ (all others)
- Transformer model is trained with **jet, electron, muon and missing transverse momentum** features ($p_x$, $p_y$, $p_z$, $p_T$, E, m, $\eta$, $\phi$, sin($\phi$), cos($\phi$))
  - Also includes DL1r score for jets and charge plus electron/muon index for leptons




Source: [HIGG-2020-24]

# ggF and VBF $HH(\to b\bar{b}b\bar{b})$ reweighting



- DNN trained for **density ratio estimation** via loss function given by

$$\mathcal{L}(w(\vec{x})) = \int d\vec{x} \left[ \sqrt{w(\vec{x})}\, p_{2b}(\vec{x}) + \frac{1}{\sqrt{w(\vec{x})}} p_{4b}(\vec{x}) \right]$$

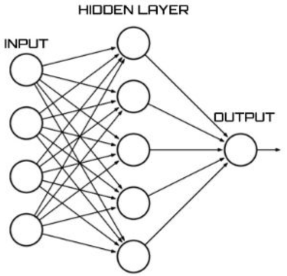
  - Minimized by **ratio of PDFs** of *4b* and *2b* samples $w(\vec{x}) = \dfrac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})}$
- **Separate trainings** for ggF and VBF (+ each data-taking year in ggF)
- Trainings are performed before splitting events by $|\Delta\eta_{HH}|$ and $X_{HH}$
  - Variables found **insensitive to kinematic reweighting**

## *Neural network inputs*

| ggF | VBF |
|---|---|
| 1. $\log(p_{\mathrm{T}})$ of the 2$^{\mathrm{nd}}$ leading Higgs boson candidate jet<br>2. $\log(p_{\mathrm{T}})$ of the 4$^{\mathrm{th}}$ leading Higgs boson candidate jet<br>3. $\log(\Delta R)$ between the closest two Higgs boson candidate jets<br>4. $\log(\Delta R)$ between the other two Higgs boson candidate jets<br>5. Average absolute $\eta$ value of the Higgs boson candidate jets<br>6. $\log(p_{\mathrm{T}})$ of the di-Higgs system<br>7. $\Delta R$ between the two Higgs boson candidates<br>8. $\Delta\phi$ between jets in the leading Higgs boson candidate<br>9. $\Delta\phi$ between jets in the subleading Higgs boson candidate<br>10. $\log(X_{Wt})$<br>11. Number of jets in the event<br>12. Trigger class index as one-hot encoder | 1. Maximum dijet mass from the possible pairings of the four Higgs boson candidate jets<br>2. Minimum dijet mass from the possible pairings of the four Higgs boson candidate jets<br>3. Energy of the leading Higgs boson candidate<br>4. Energy of the subleading Higgs boson candidate<br>5. Second-smallest $\Delta R$ between the jets in the leading Higgs boson candidate (from the three possible pairings for the leading Higgs candidate)<br>6. Average absolute $\eta$ value of the four Higgs boson candidate jets<br>7. $\log(X_{Wt})$<br>8. Trigger class index as one-hot encoder<br>9. Year index as one-hot encoder (for years inclusive training) |

Source: HDBS-2019-29

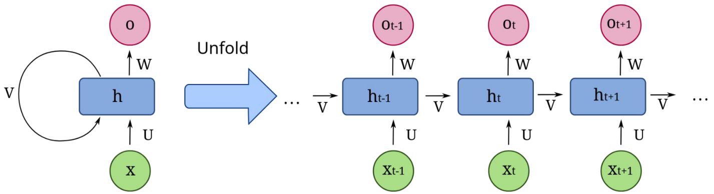
ML architecture: depends on the structure of the data and the target task



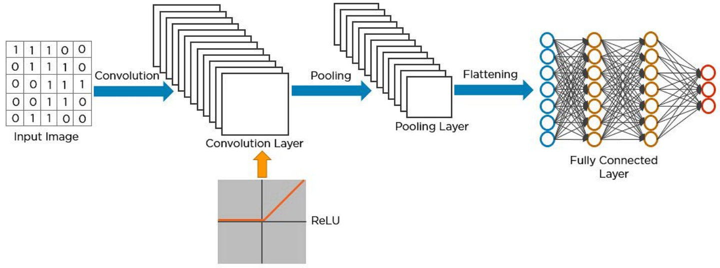

**Feed-forward NN**

Universal function approximator.

No encoded invariance in the input




**Recurrent NN:** used for sequences, can encode casualty,
Inefficient for long sequences



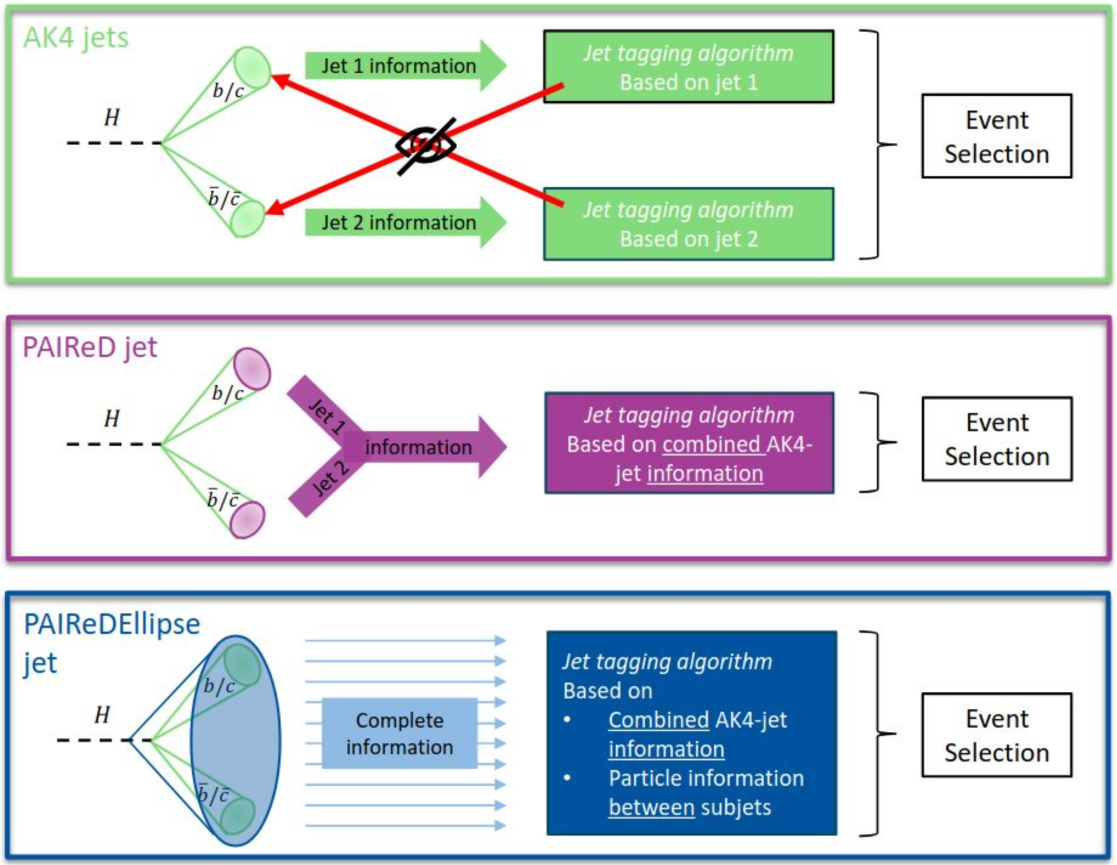

**Convolutional NN:** used for images, inefficient for sparse data.
Encodes input invariance by translation

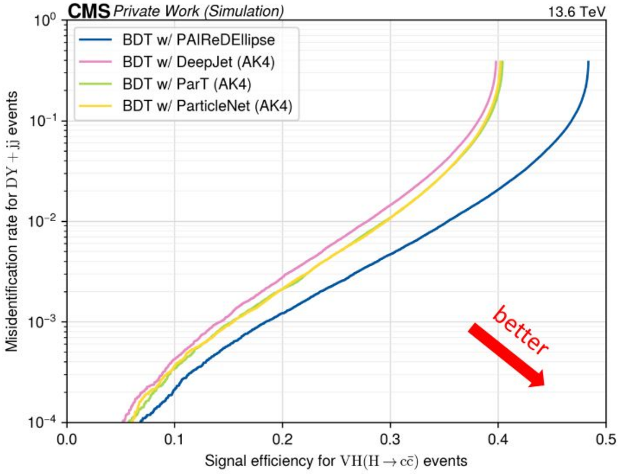
Have a look at the Living Review of Machine Learning for Particle Physics for an updated list of ML in HEP applications

# PAIRed-jet classifier

ETH zürich

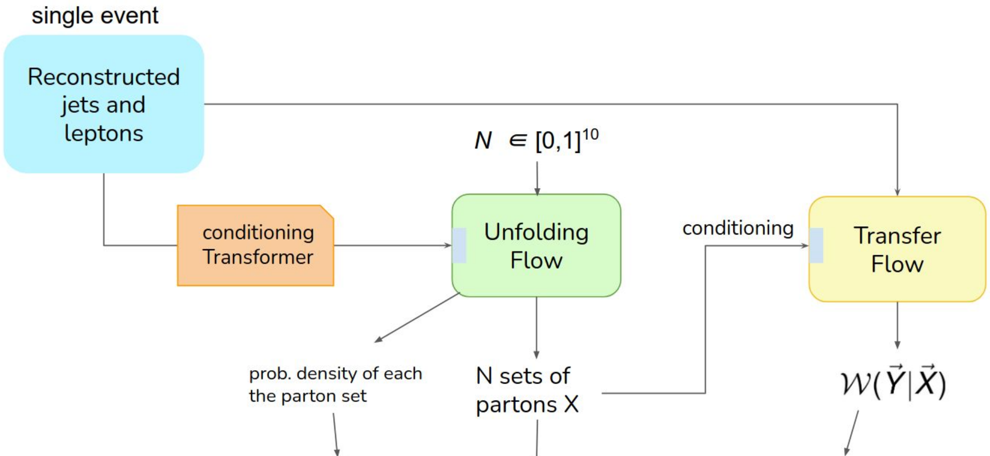
[arXiv:2311.11011 [hep-ex]; accepted by JHEP]



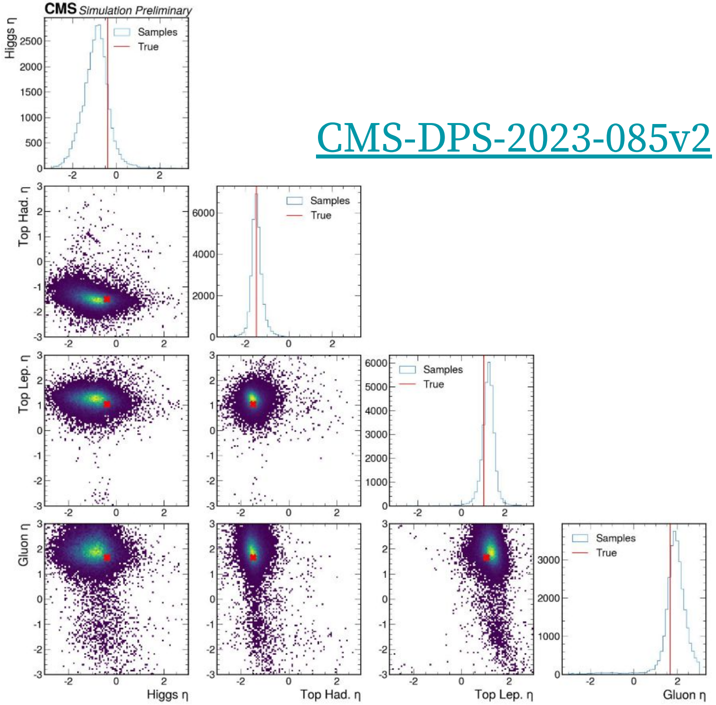



# Generative ML for MEM

Computing the Matrix Element Method for ttH(bb) using Normalizing Flows and transformers for optimal sampling and *P(Reco|Gen)* transfer functions




$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$




CMS-DPS-2023-085v2