Transformer networks for constituent-based b-jet calibration with the ATLAS detector

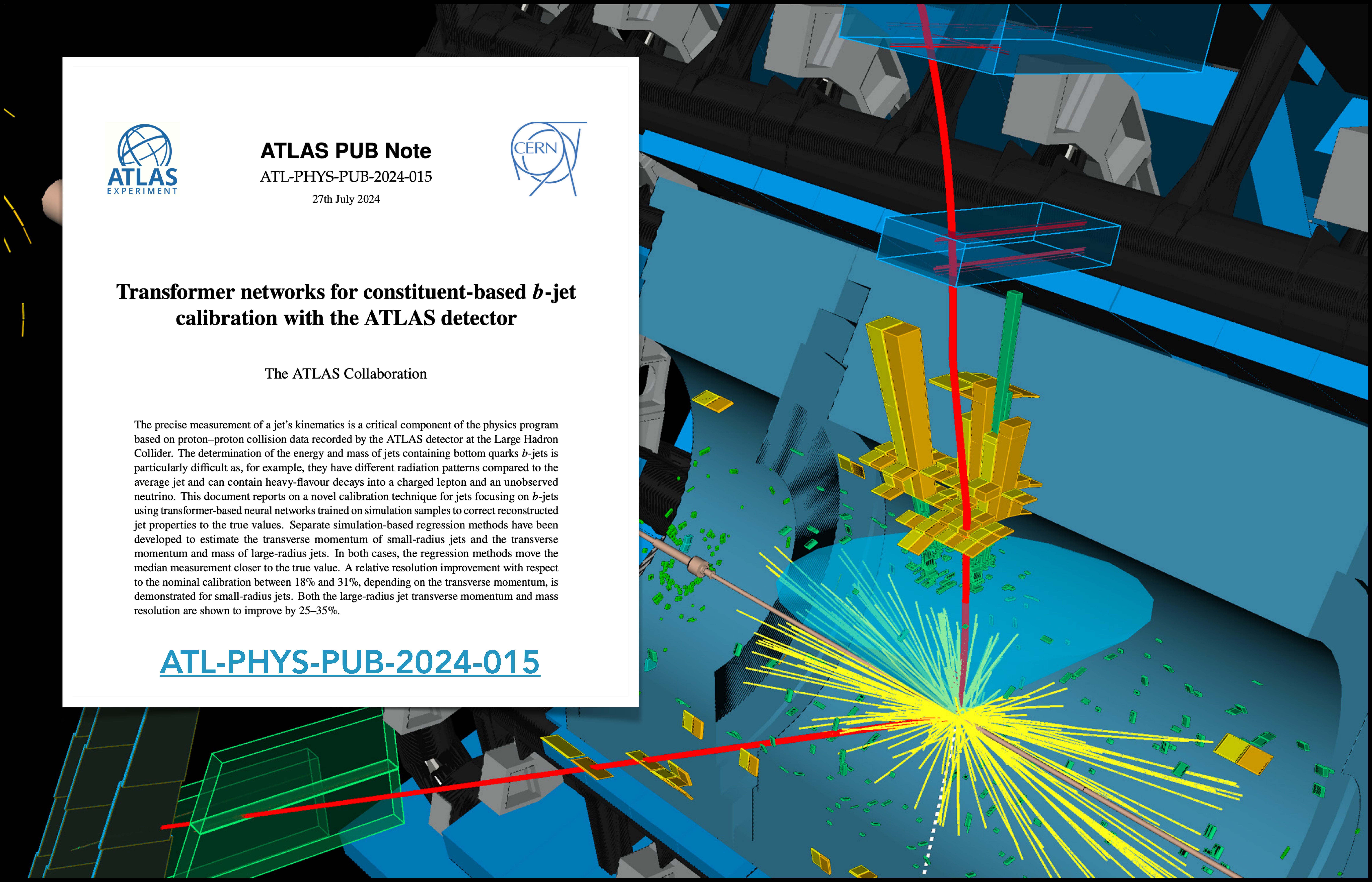
Brendon Bullard
on behalf of the ATLAS Collaboration

SLAC National Accelerator Laboratory

ML4Jets - Reconstruction Session
November 5, 2024

ATLAS EXPERIMENT

SLAC NATIONAL ACCELERATOR LABORATORY



ATLAS PUB Note

ATL-PHYS-PUB-2024-015

27th July 2024

Transformer networks for constituent-based b-jet calibration with the ATLAS detector

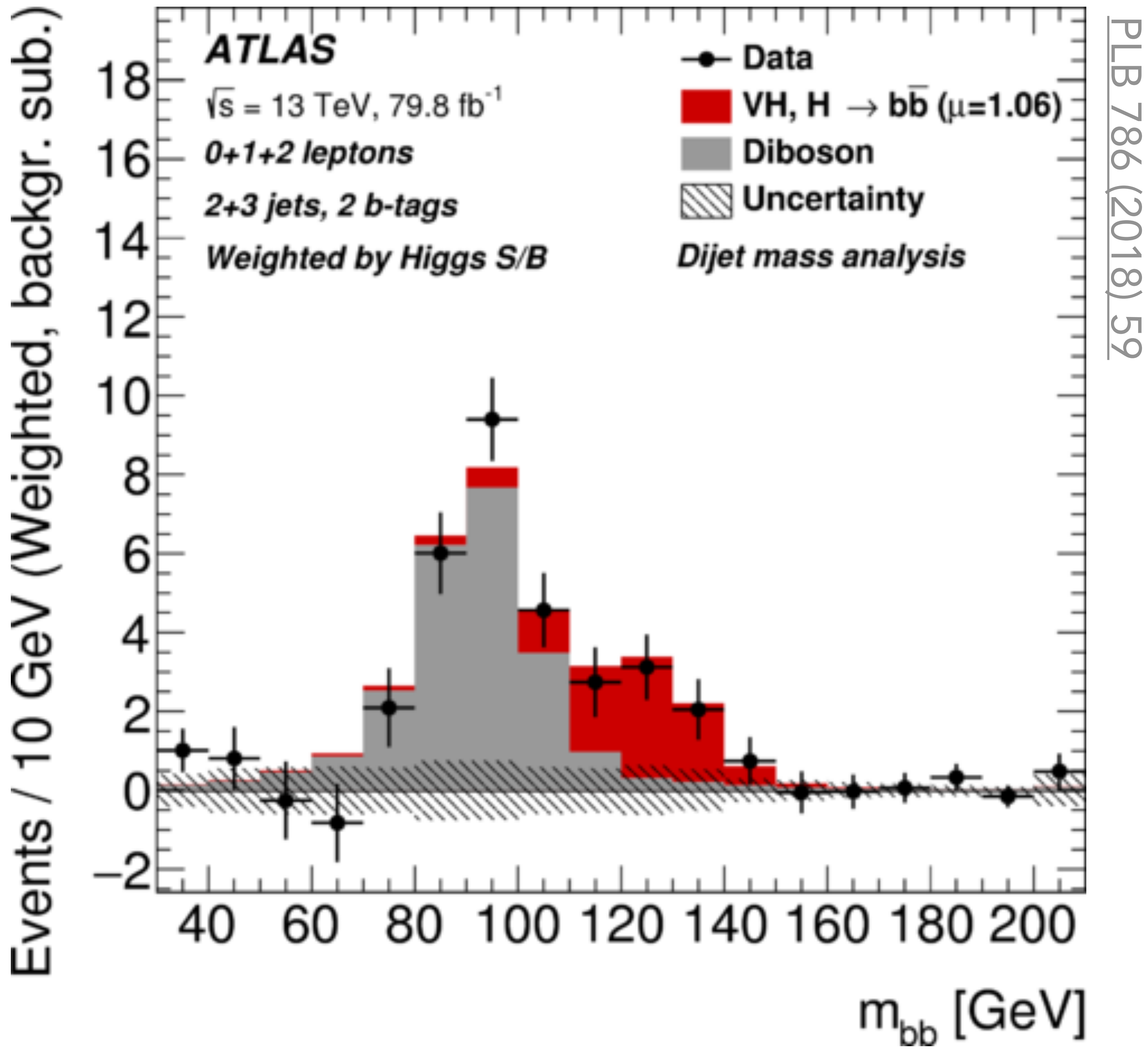
The ATLAS Collaboration

The precise measurement of a jet's kinematics is a critical component of the physics program based on proton–proton collision data recorded by the ATLAS detector at the Large Hadron Collider. The determination of the energy and mass of jets containing bottom quarks b-jets is particularly difficult as, for example, they have different radiation patterns compared to the average jet and can contain heavy-flavour decays into a charged lepton and an unobserved neutrino. This document reports on a novel calibration technique for jets focusing on b-jets using transformer-based neural networks trained on simulation samples to correct reconstructed jet properties to the true values. Separate simulation-based regression methods have been developed to estimate the transverse momentum of small-radius jets and the transverse momentum and mass of large-radius jets. In both cases, the regression methods move the median measurement closer to the true value. A relative resolution improvement with respect to the nominal calibration between 18% and 31%, depending on the transverse momentum, is demonstrated for small-radius jets. Both the large-radius jet transverse momentum and mass resolution are shown to improve by 25–35%.

ATL-PHYS-PUB-2024-015

Physics motivation

Measuring $H \rightarrow b\bar{b}$ constrains **bottom Yukawa**



Limited by poor **jet momentum resolution** and large continuum multi-jet and $Z \rightarrow b\bar{b}$ bkg

Di-Higgs production is a critical target, gives handle on **Higgs self-coupling**

	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

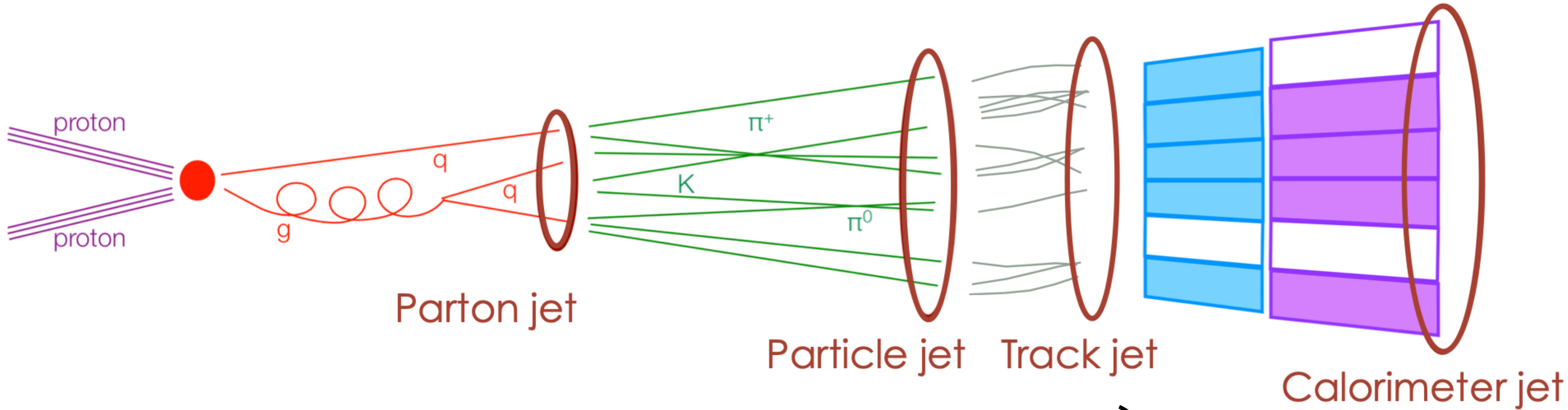
70%!

Branching fractions of the the two Higgs

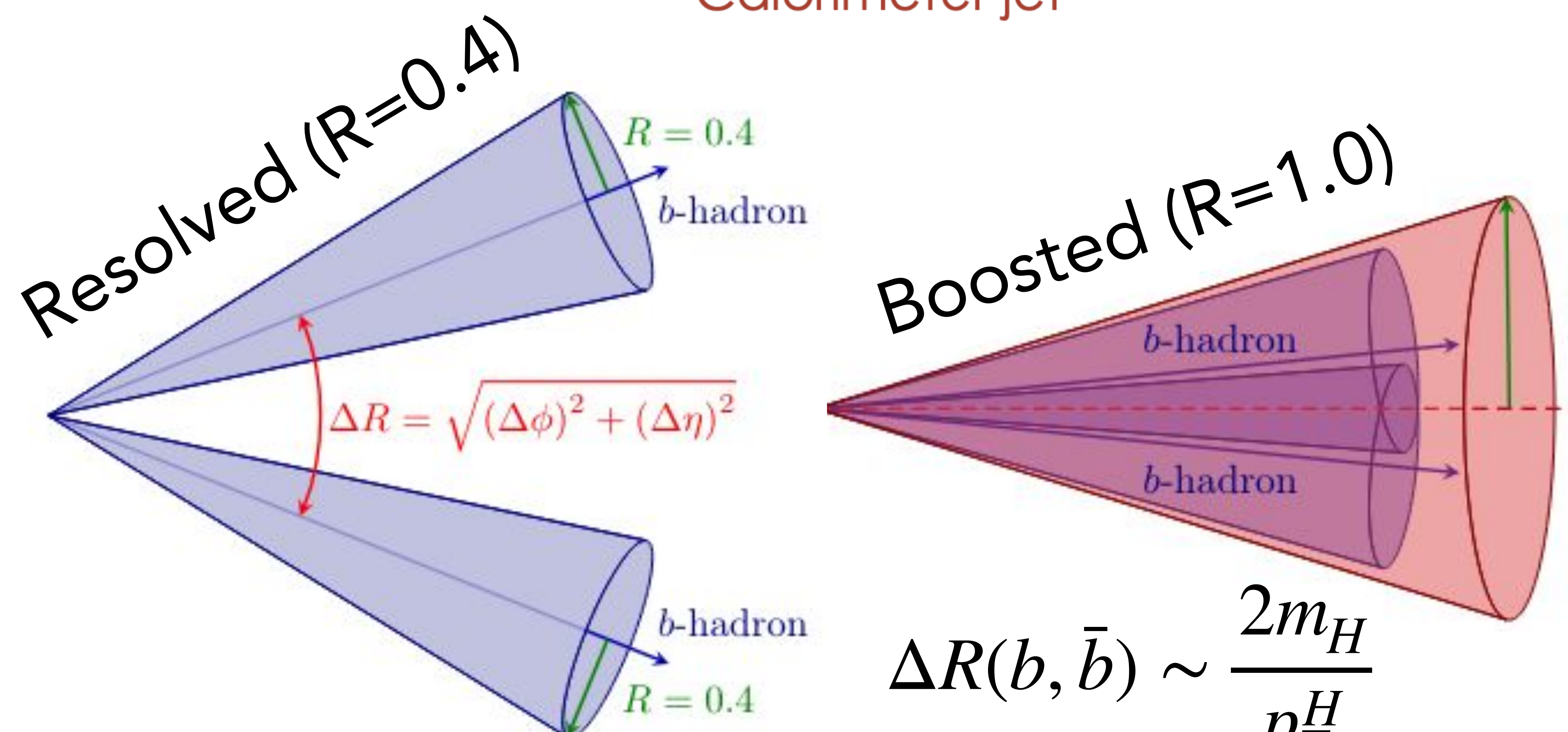
Improving the **reconstruction of b-quark jets** has huge impact

Jet reconstruction

- Jets are the most complex objects produced at colliders
  - Needs careful combination of signatures in tracker and calorimeters



- Cluster constituents using anti-k_T algorithm
  - Use different radius parameters depending on the target phase space (e.g. low/high-p_T Higgs)

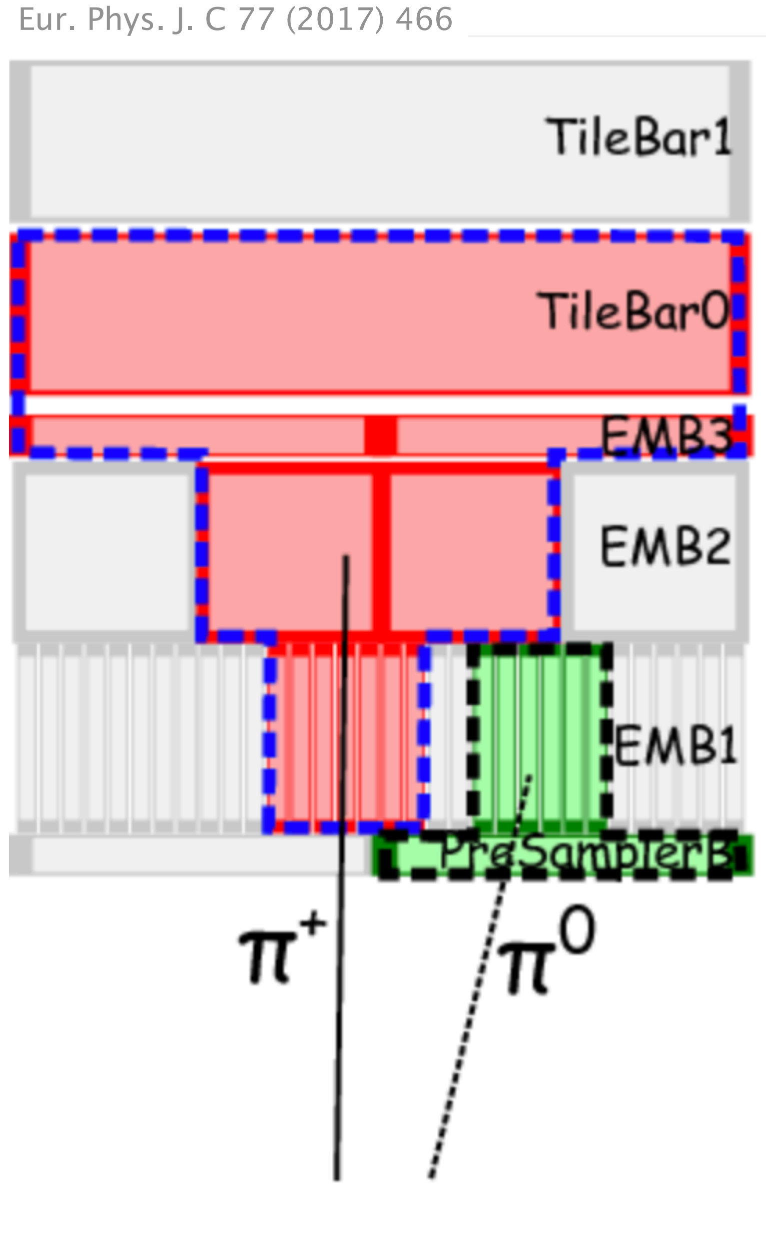


$$\Delta R(b, \bar{b}) \sim \frac{2m_H}{p_T^H}$$

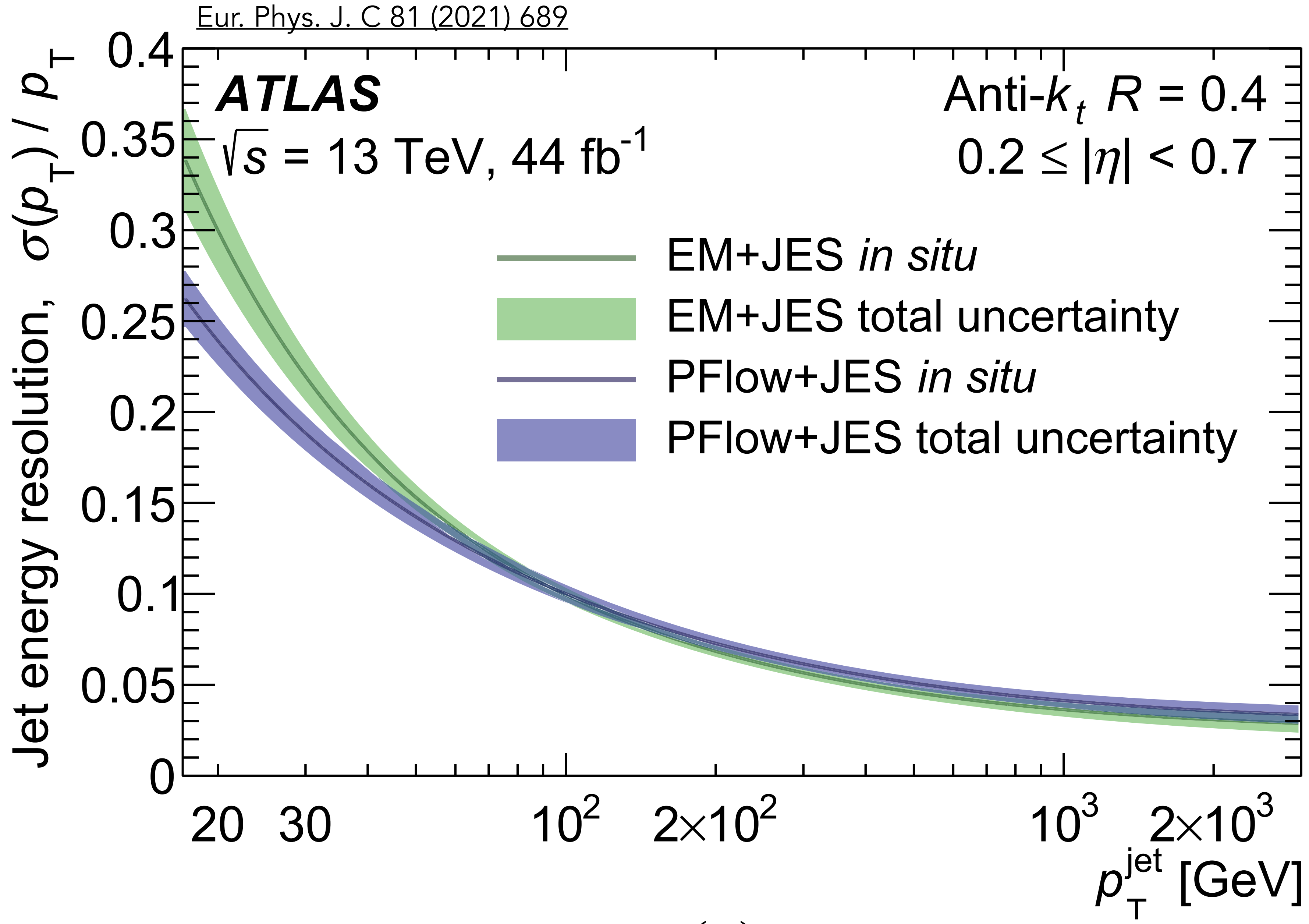
Jet constituents

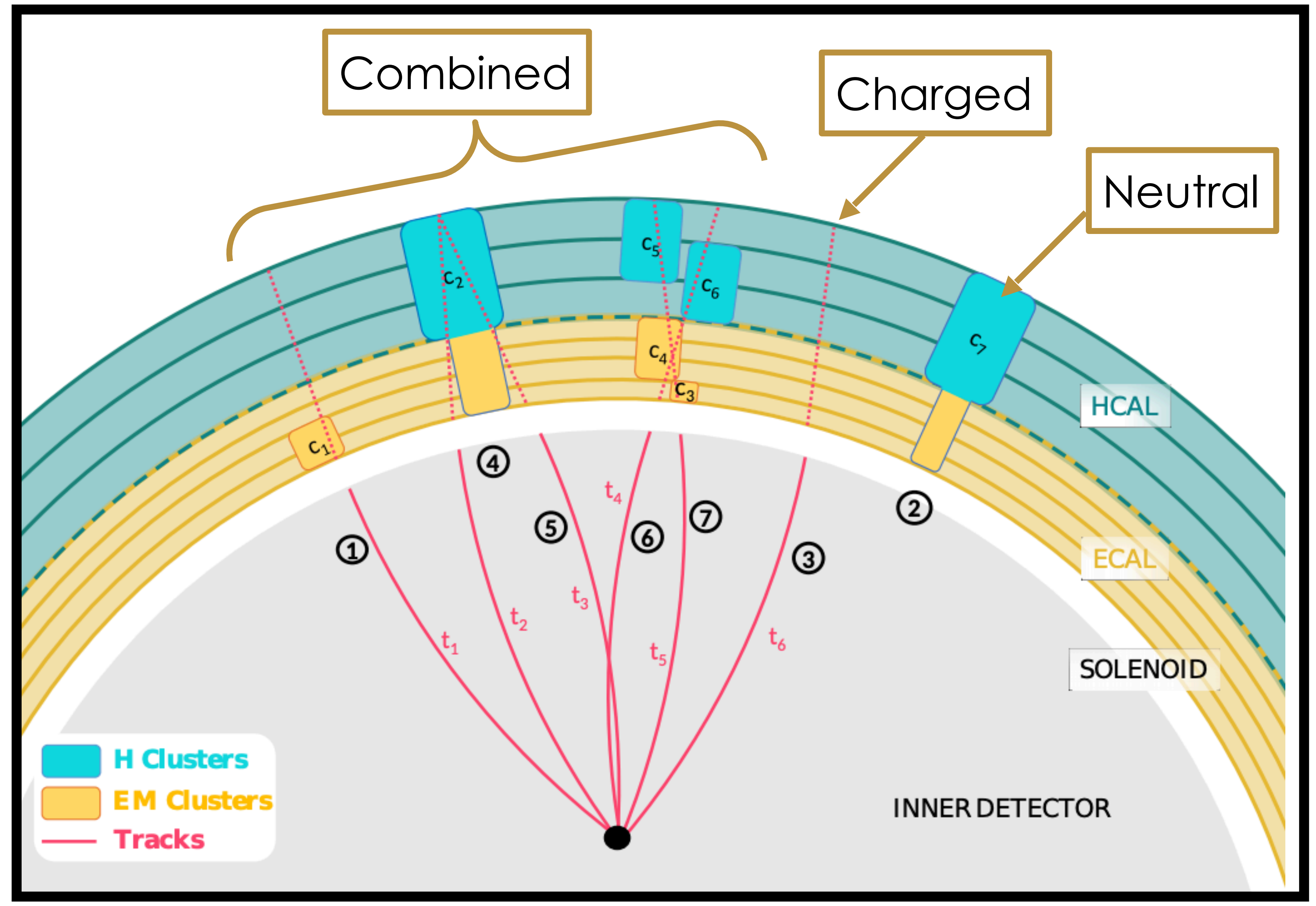
- Originally ATLAS only used calorimeter cells for jets
- Now we are using **particle flow** objects: combine tracks and calo-clusters
  - Avoid double-counting energy/momentum, boosts performance at low-p_T (used for Small-R)
- Recently developed **unified flow** objects, leverage angular resolution of tracker and energy resolution of calorimeter at high-p_T (used for Large-R jets)

Eur. Phys. J. C 77 (2017) 466



Eur. Phys. J. C 81 (2021) 689



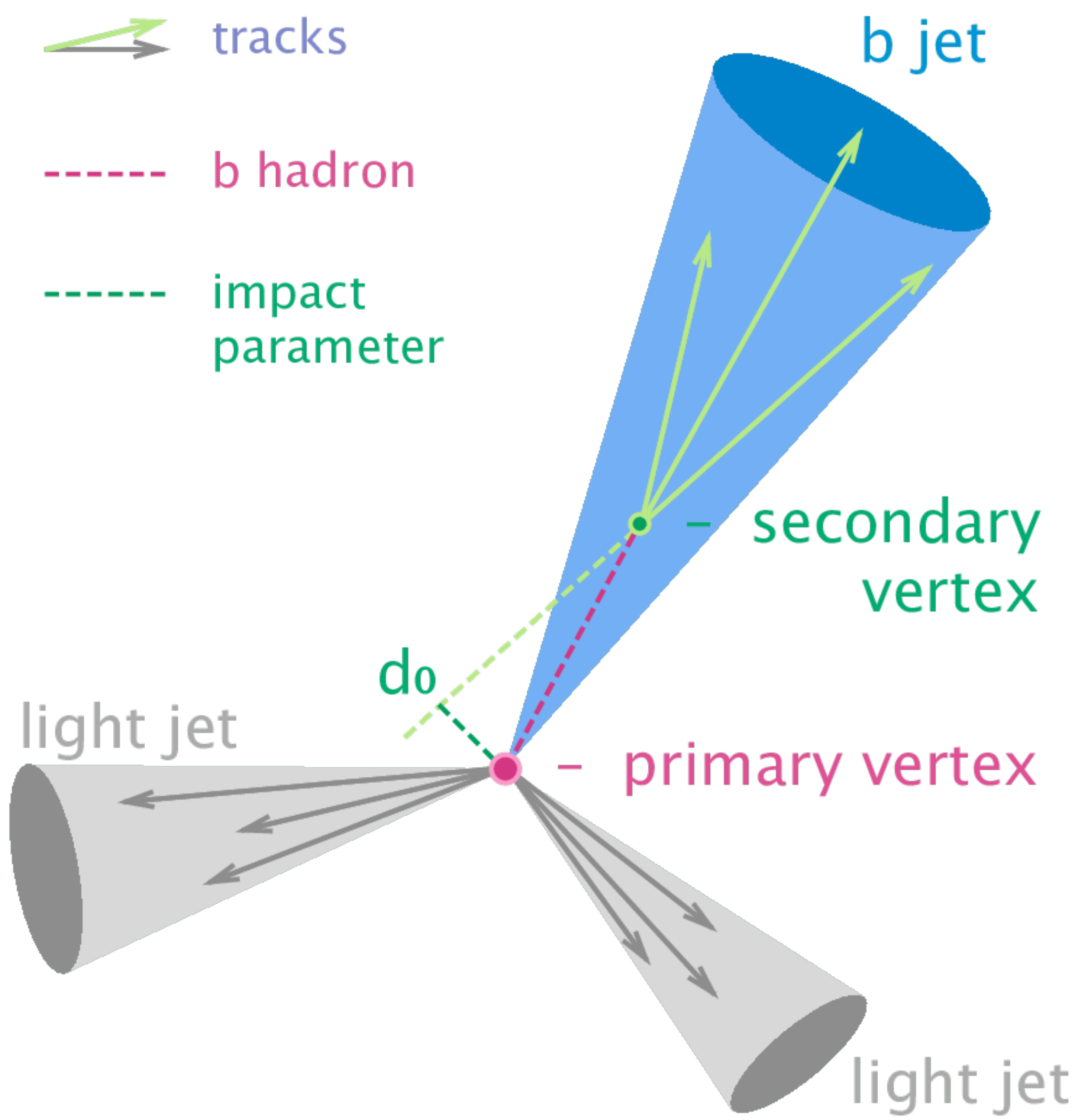


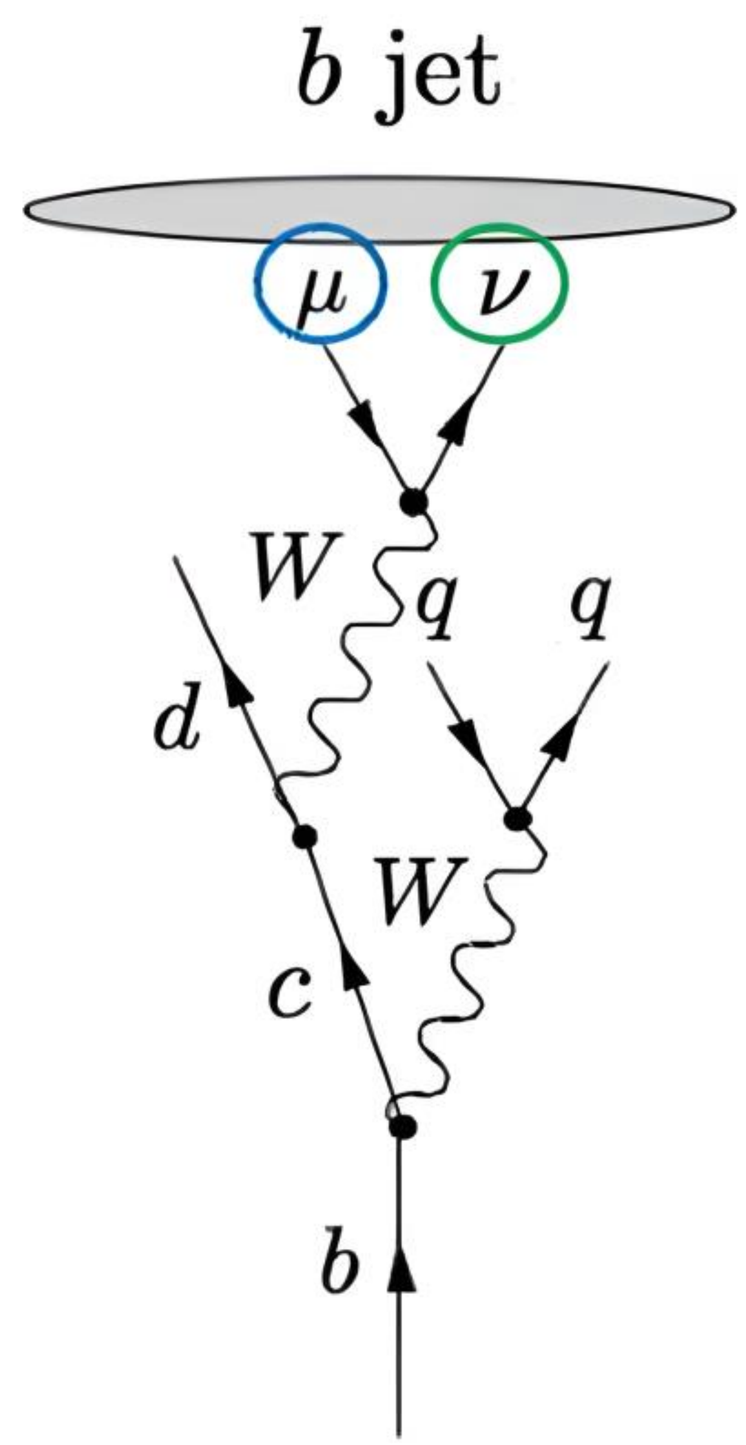
ATL-PHYS-PUB-2017-15

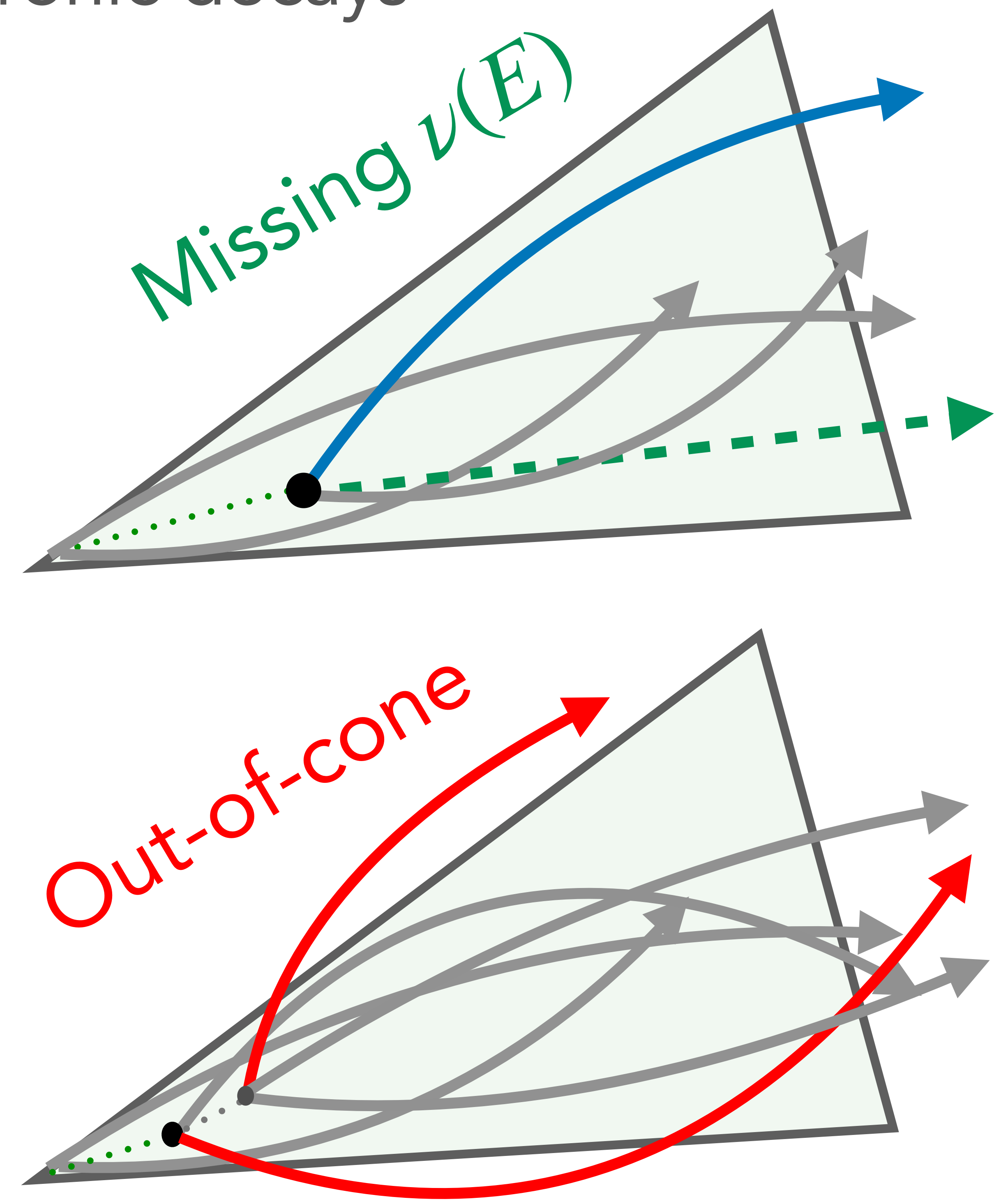
Physics of b-jets

SLAC

- ✦ B-jet signatures are unique due to secondary vertex
  - The B-hadron carries >80% of the jet energy
  - Semi-leptonic decays (~15% BR), ν carries 20% of the jet energy
- ✦ Baseline strategy: **μ-in-jet** addition to jet 4-mom, apply **PtReco correction** to jet p_T
  - Coarse-grained correction binned in jet p_T, split into leptonic and hadronic decays



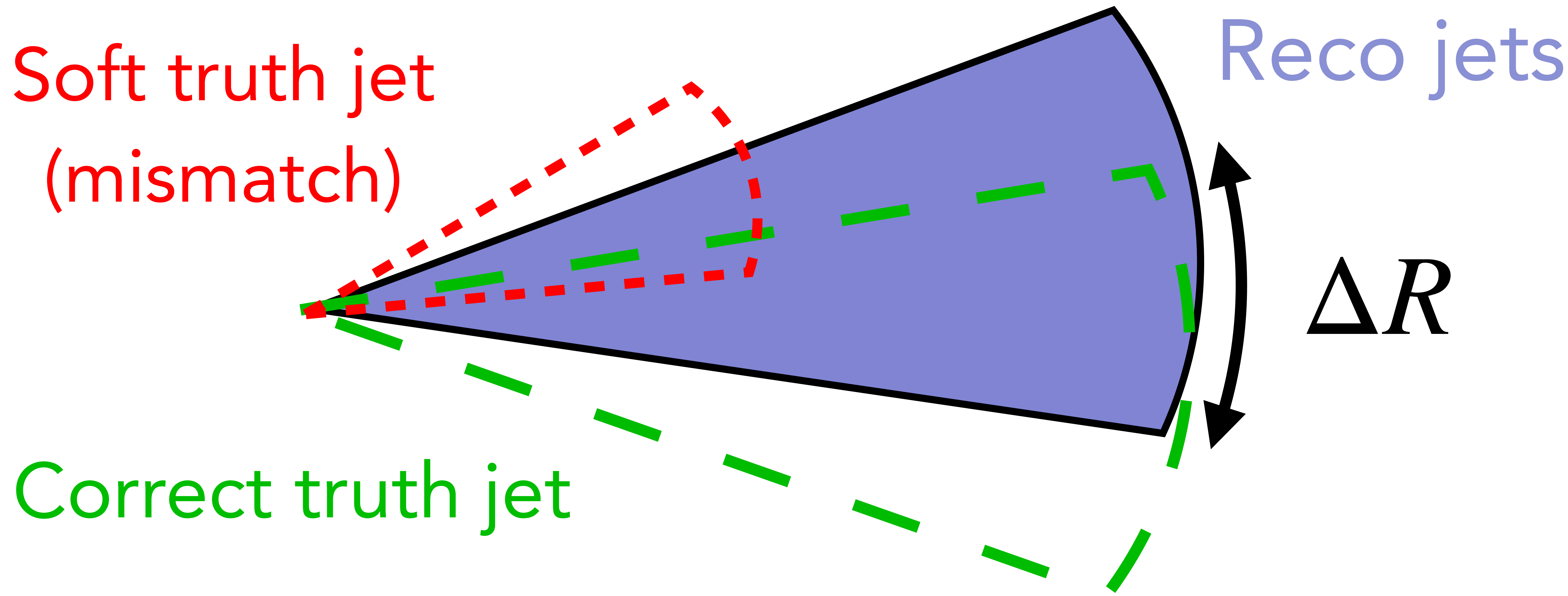




Jet definitions

- Define jets at truth-level and match to reco-level using ΔR matching
  - Small-R: include neutrinos and muons in truth jet definition
  - Large-R: not including neutrinos and muons, only correcting hadronic activity (for now)
- Low-p_T/mass thresholds for jets to avoid bias in the training
  - Small-R: 7 (10) GeV for truth (reco)-level jets — will calibrate $p_T > 20$ GeV
  - Large-R: $200 < p_T < 1500$ GeV, jet mass between [20, 300] GeV

Select truth jet with largest p_T within $\Delta R < 0.4$ (0.75) → robust to presence of nearby soft radiation

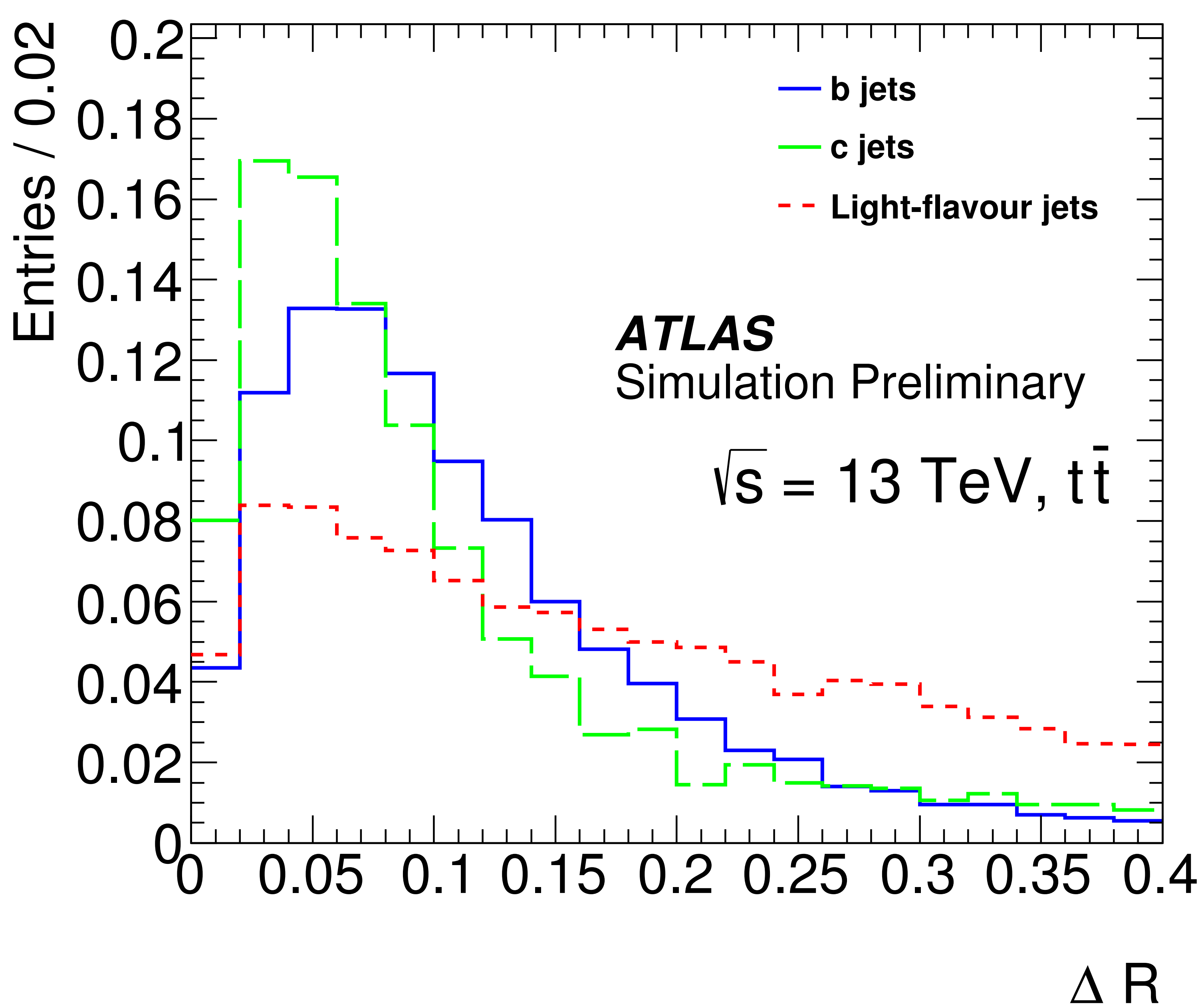


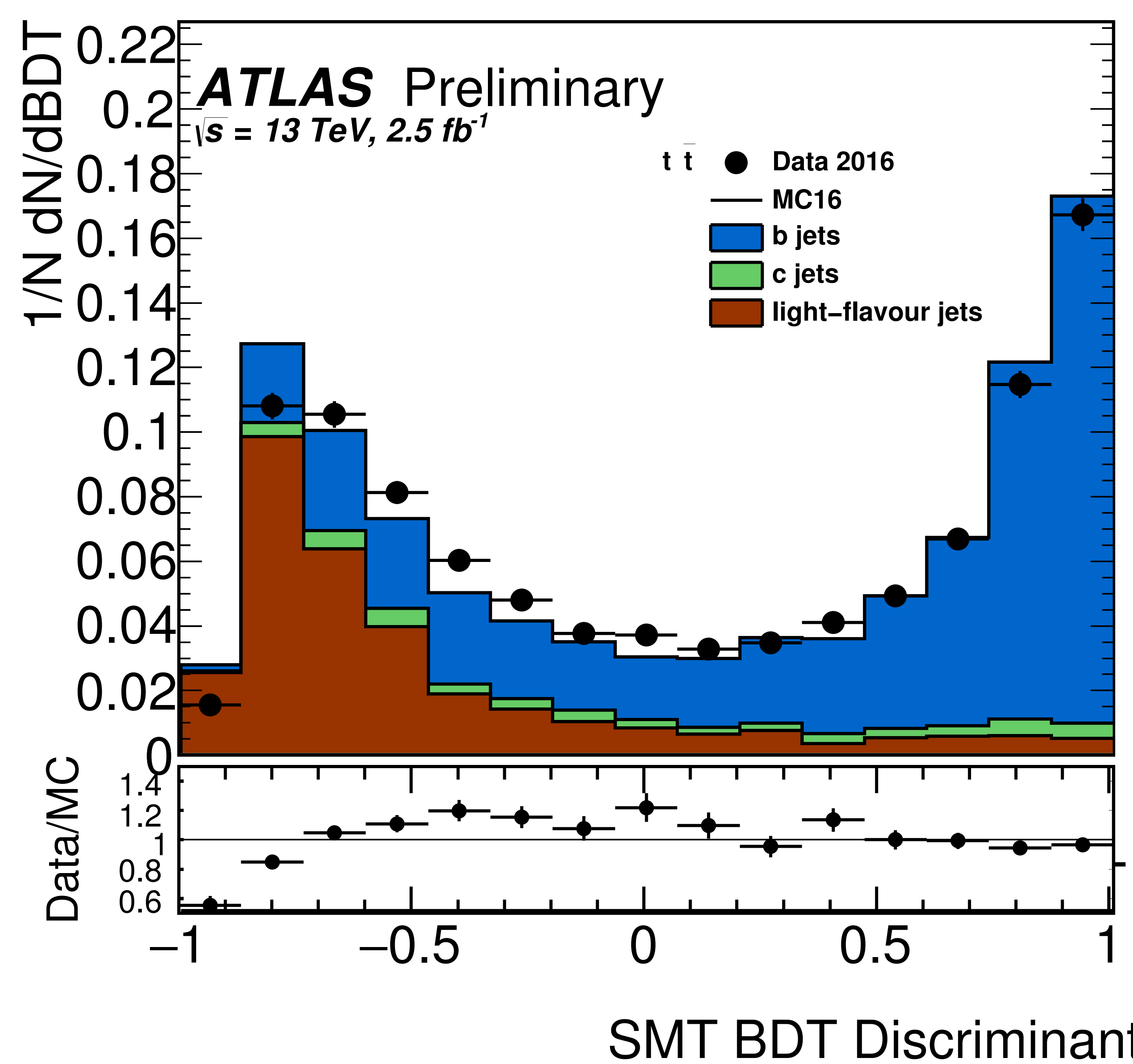
Constituent definitions

- Apply tight selection requirements for tracks and charged flow
- Use soft muon and soft electron taggers for small-R jets
  - Boosted decision trees using lepton kinematic variables relative to the jet
  - Enhance sensitivity to semi-leptonic decays of B-hadrons

EPS-HEP-2017

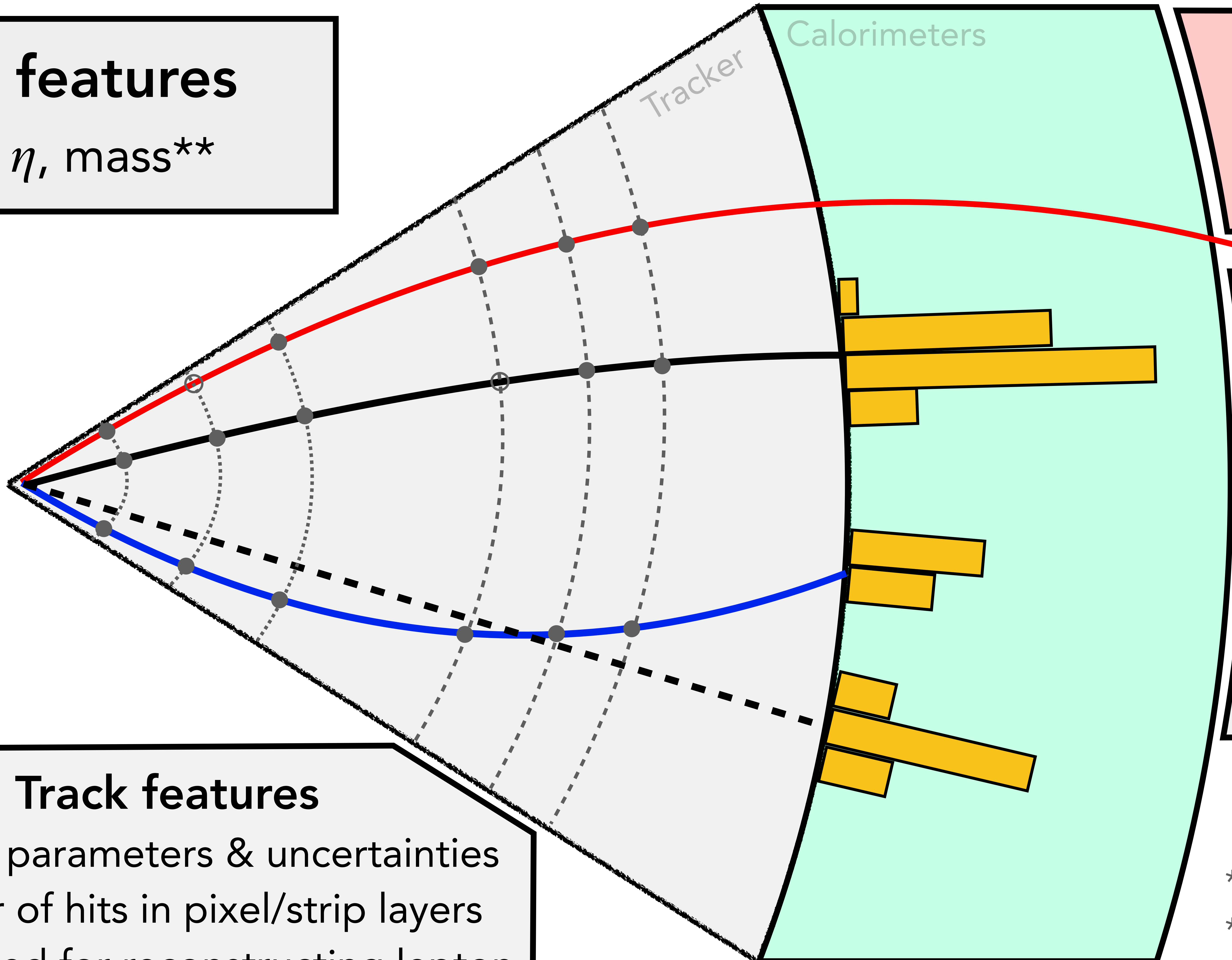
Parameter	Requirement
Track Selection	
p_T	> 500 MeV
Silicon hits	≥ 8
Shared silicon hits	≤ 1
Silicon holes	< 2
Pixel holes	< 1
Track-to-Vertex Association	
$\|d_0\|$	< 3.5 mm
$\|z_0 \sin\theta\|$	< 5 mm





Neural network input features

Jet features
p_T, η, mass**



Soft muon features*
Kinematics, perigee, p_T balance in ID/MS,

Soft electron features*
Kinematics, perigee, E/p, E_{ECal}/E_{HCal}, layer cluster ratios, p_{HF}

Flow object features**
p_T, E_T, $d\eta$, $d\phi$, dR

Track features
- Perigee parameters & uncertainties
- Number of hits in pixel/strip layers
- Track used for reconstructing lepton

* = used in small-R jets only
** = used in large-R jets only
Detailed list in backup

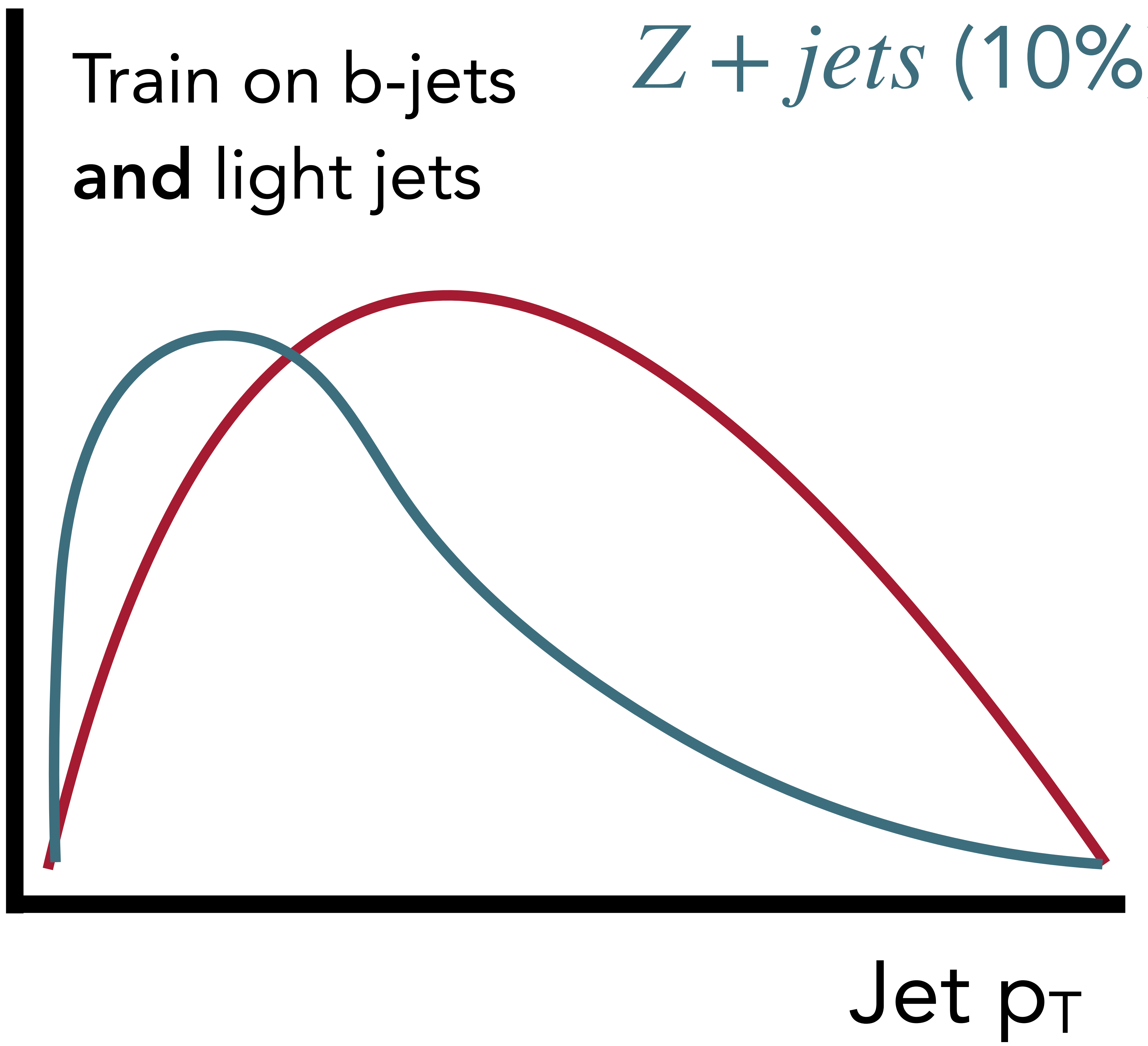
Training samples

Small-R jets

~260 million jets

$t\bar{t}$ (90%)

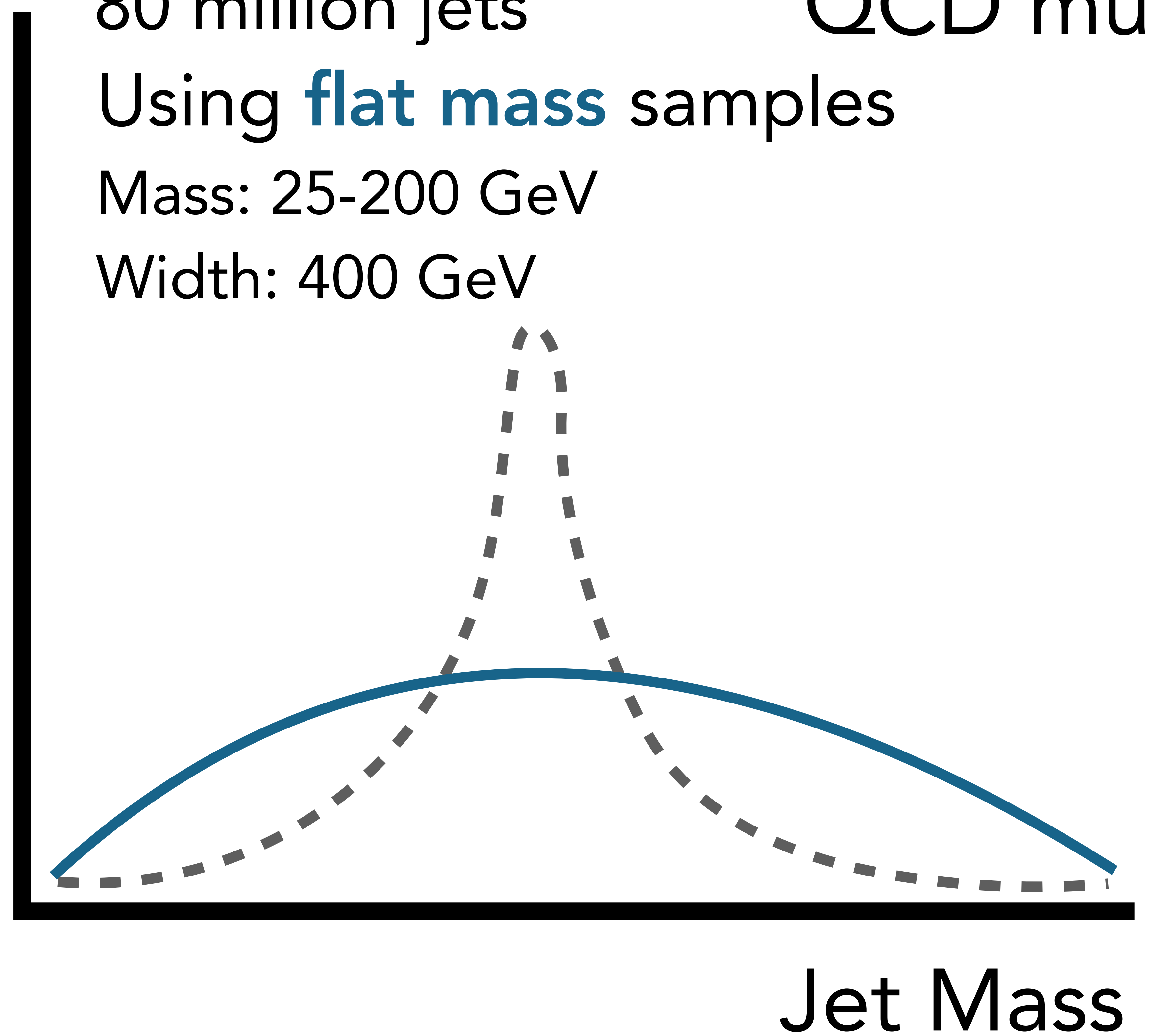
$Z + jets$ (10%)



Large-R jets

H→qq/cc/bb

QCD multijet



Details in backup

Neural network architecture

- Based on ATLAS flavor-tagging architecture
  - anti-k_T R=0.4 PFlow (small-R) jets use **track** constituents
  - anti-k_T R=1.0 UFO (large-R) jets use **track** and **flow object** constituents

Constituent embedding

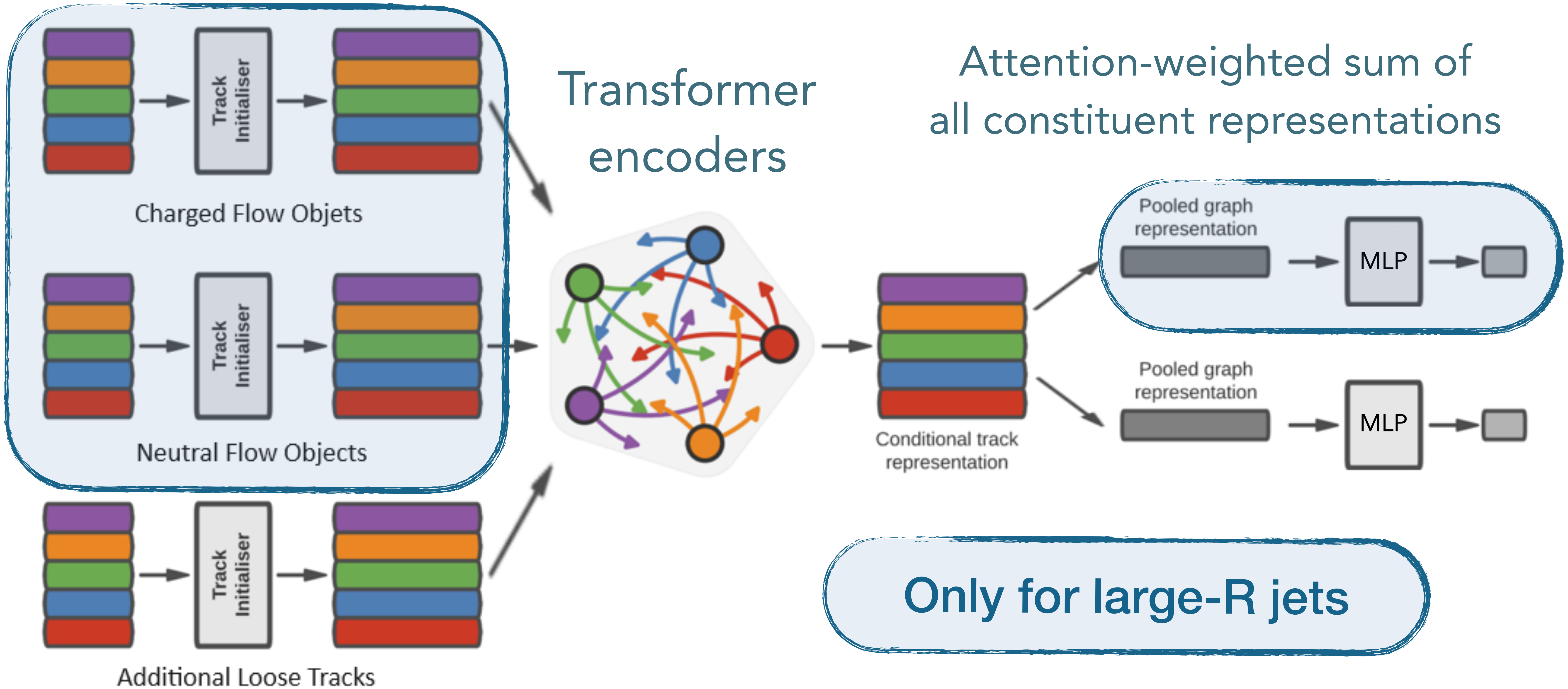
Transformer encoders

Attention-weighted sum of all constituent representations

Regression targets

Jet mass

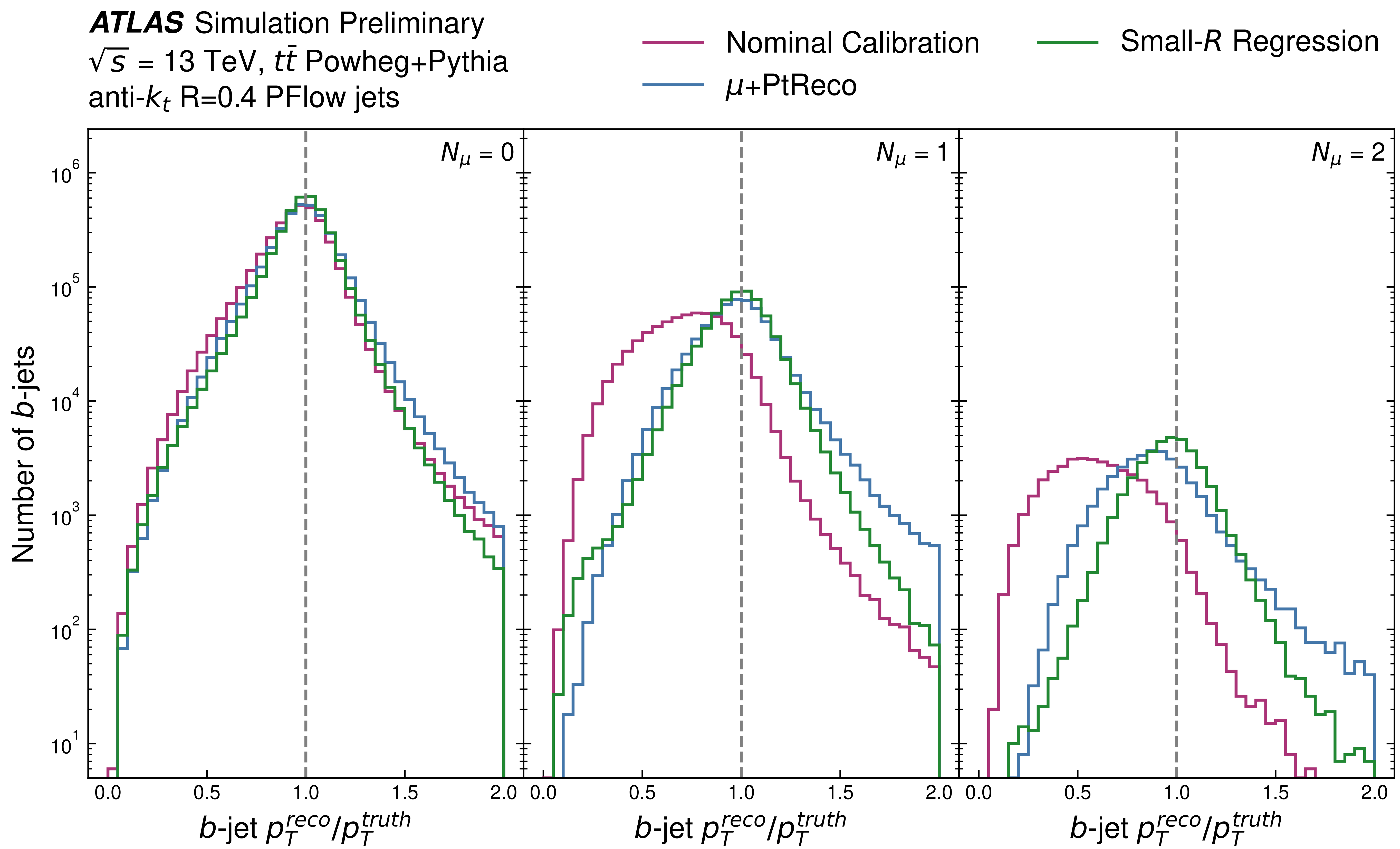
Jet p_T



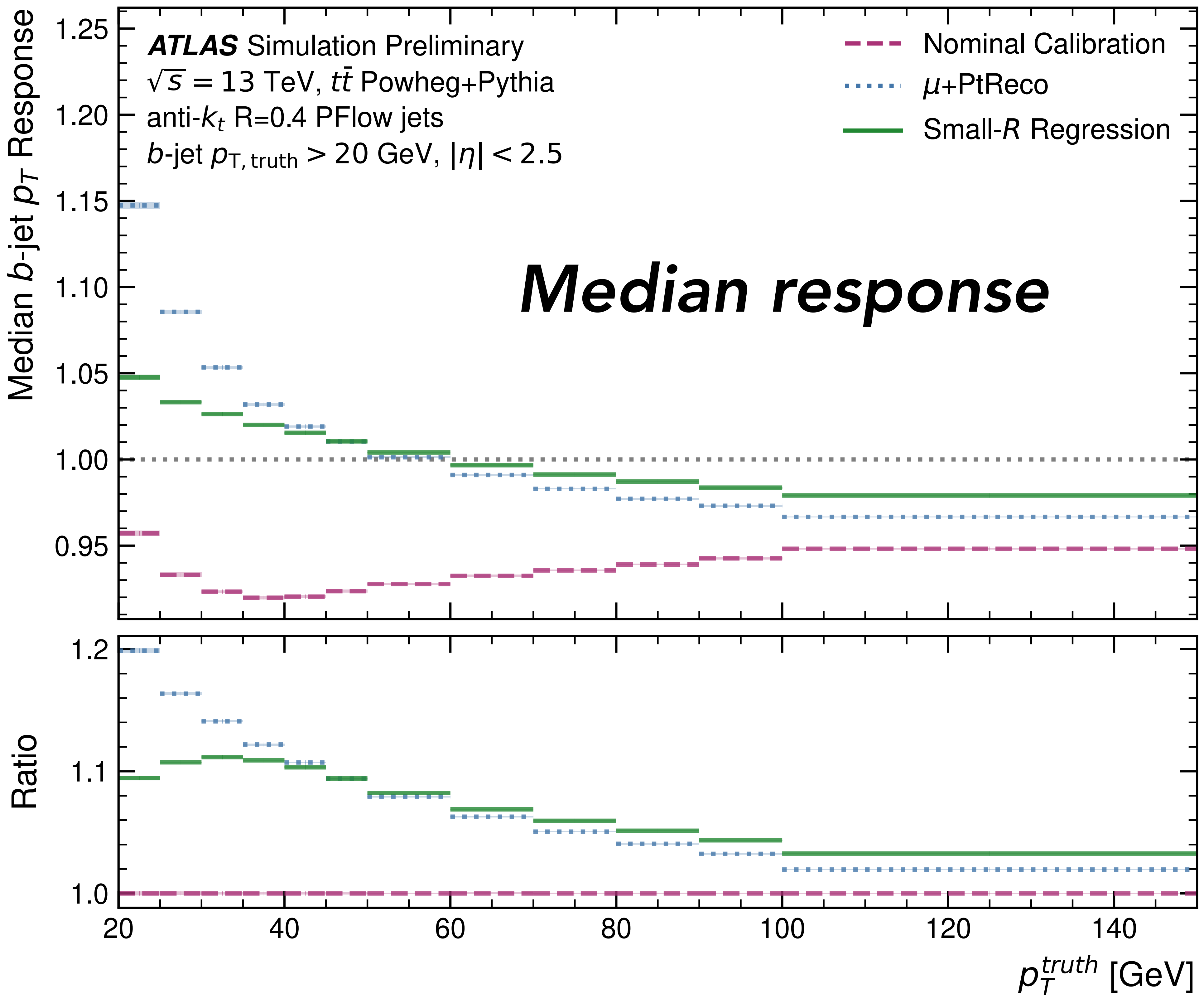
Only for large-R jets

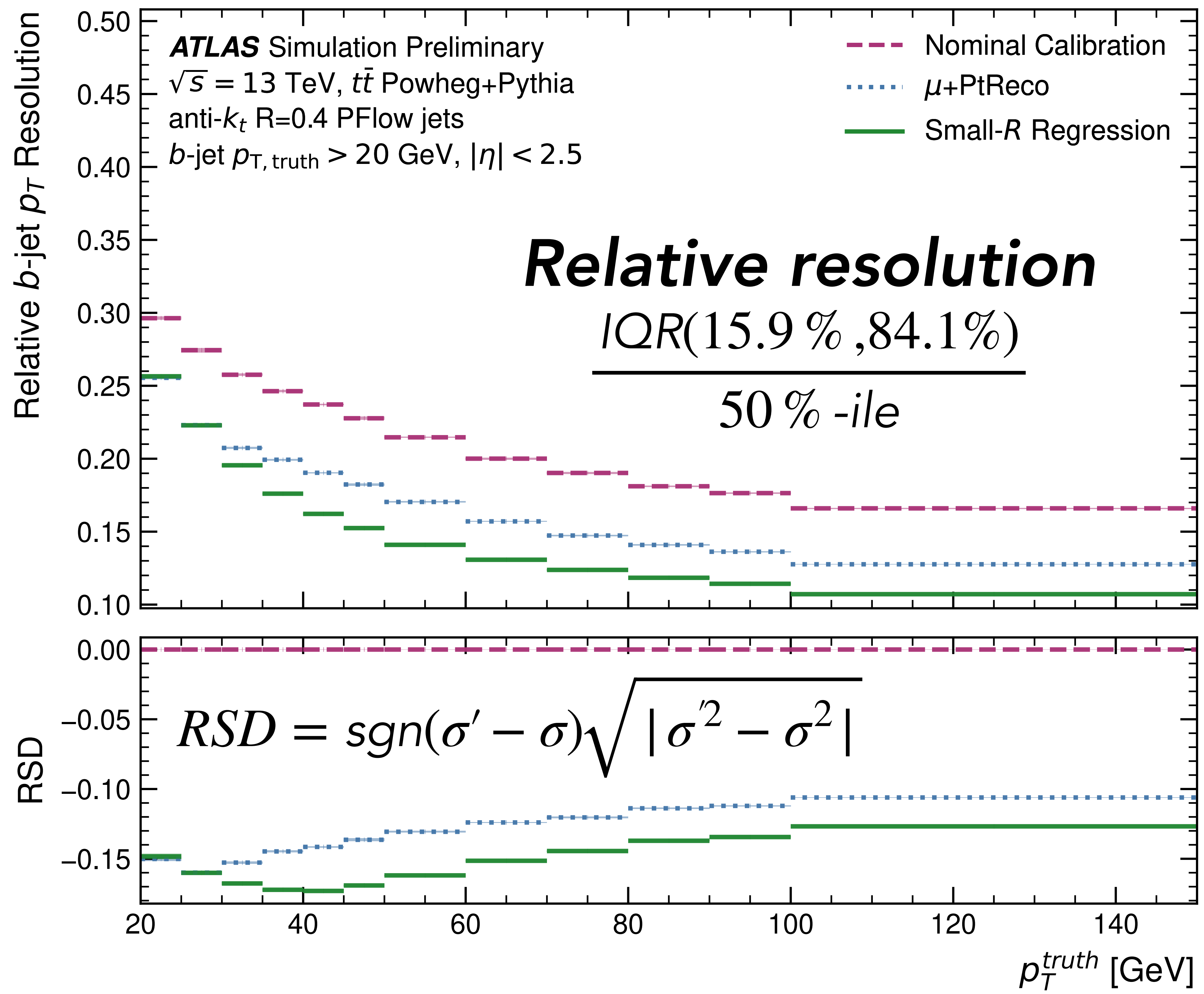
Response distributions

- Key observable is p_T (and mass) response
  - Defined by p_T^{reco}/p_T^{truth}
- Study NN performance as a function of N_{muons}
  - Proxy for B-decay channel
- Looking for narrow peaks in the response distribution around 1.0



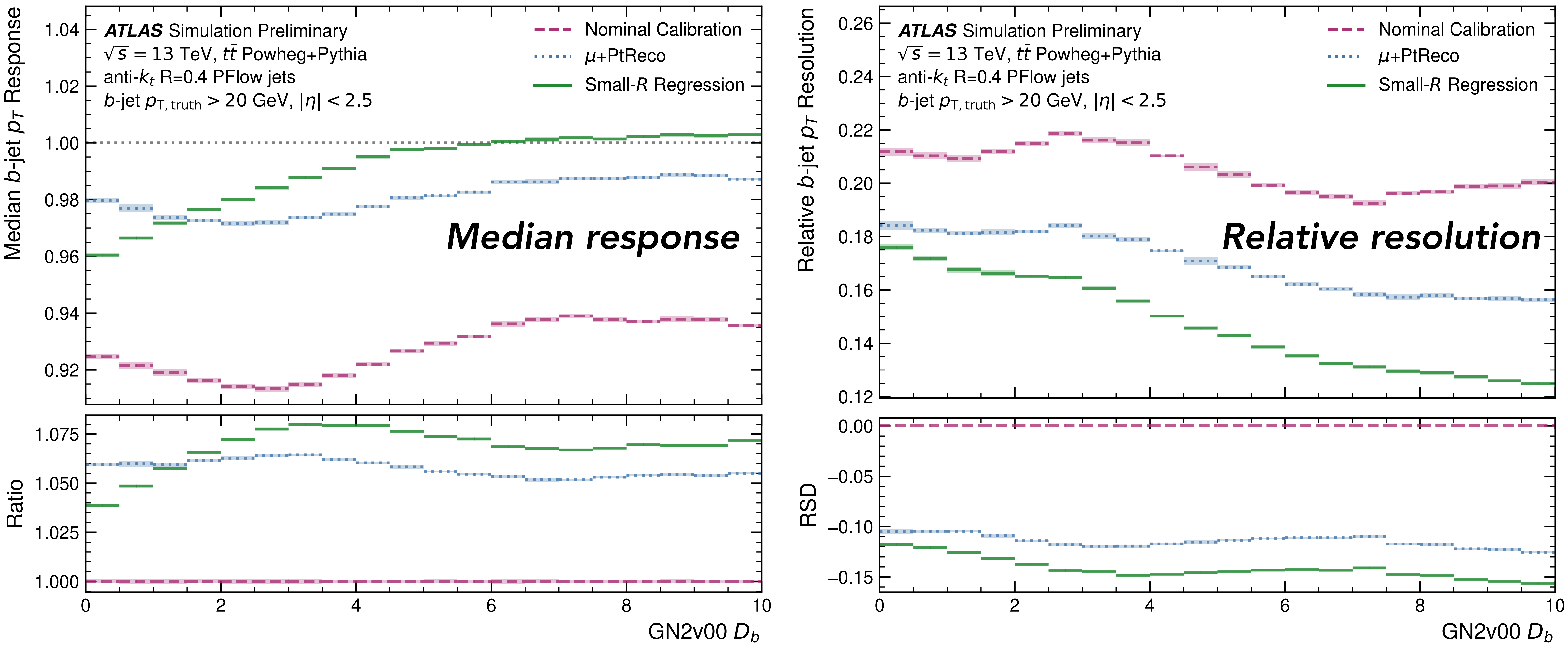
Small-R performance vs. jet p_T





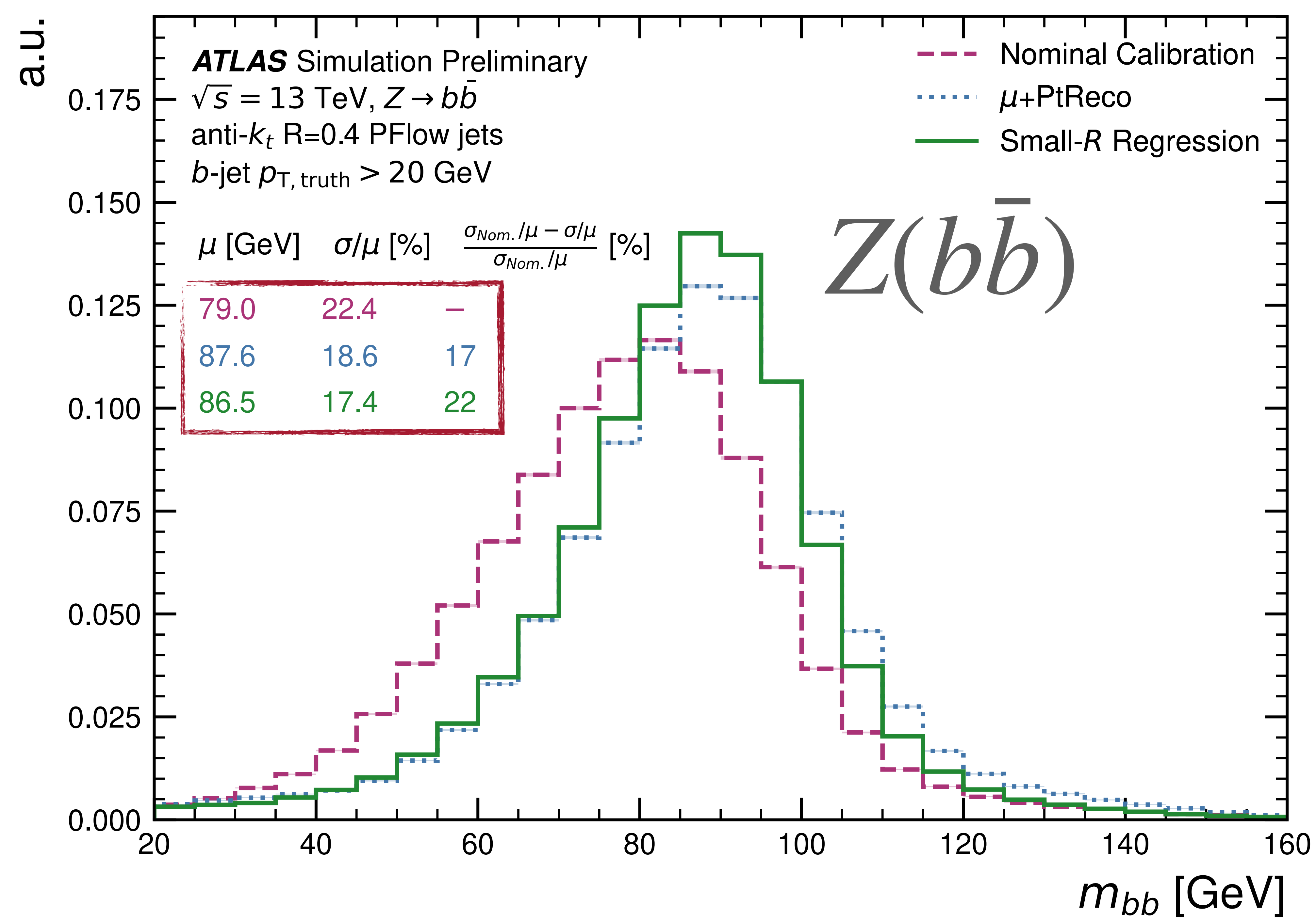
- Median response flattens with regression
- Relative resolution improves up to 30%, only 20% for PtReco

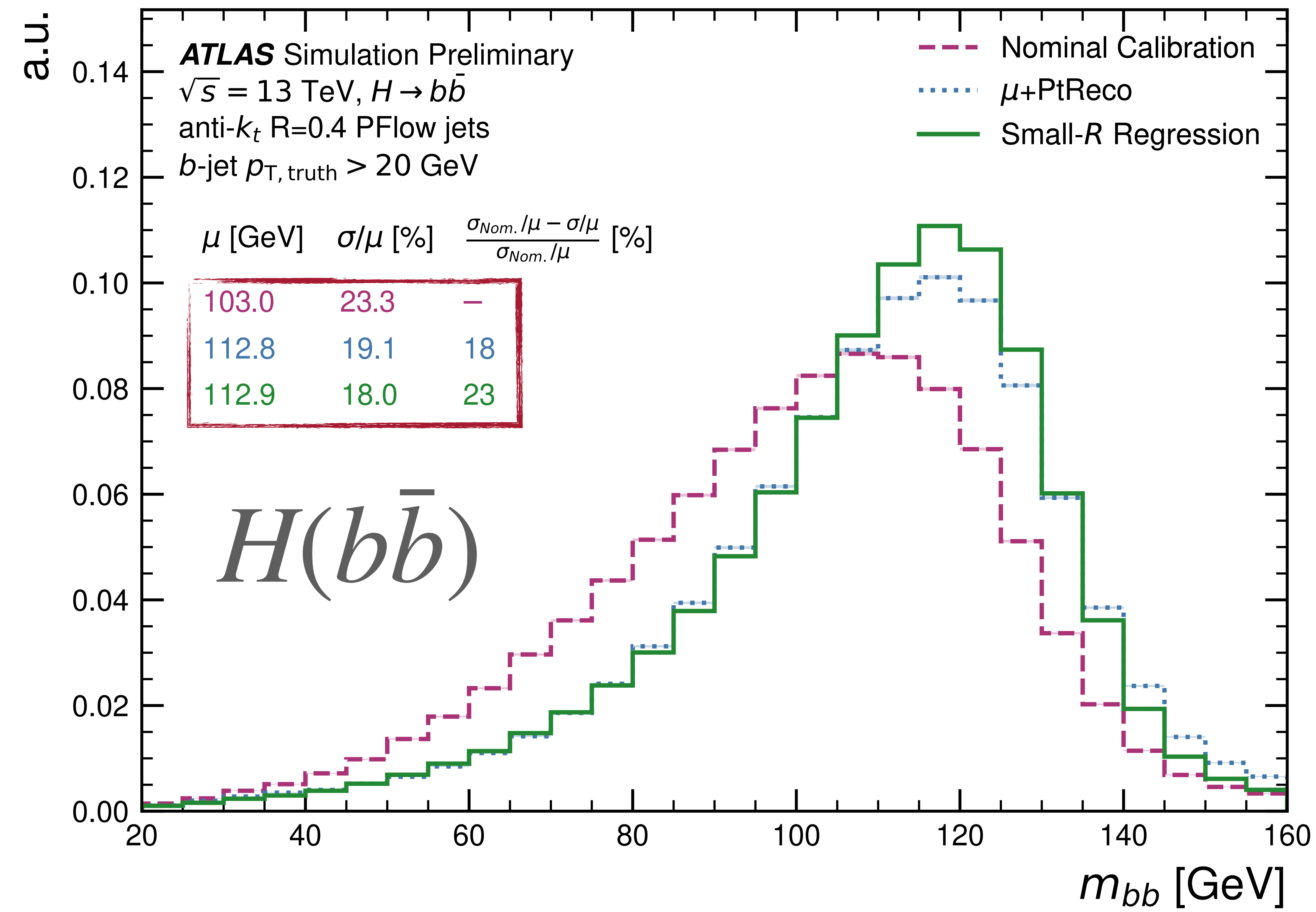
Small-R performance vs. flavor tag



- Even though they are independent algorithms, clear correlation between flavor-tagging discriminant and regression performance
  - Honing in on the **same signatures** — B-decay length, track multiplicity, etc.

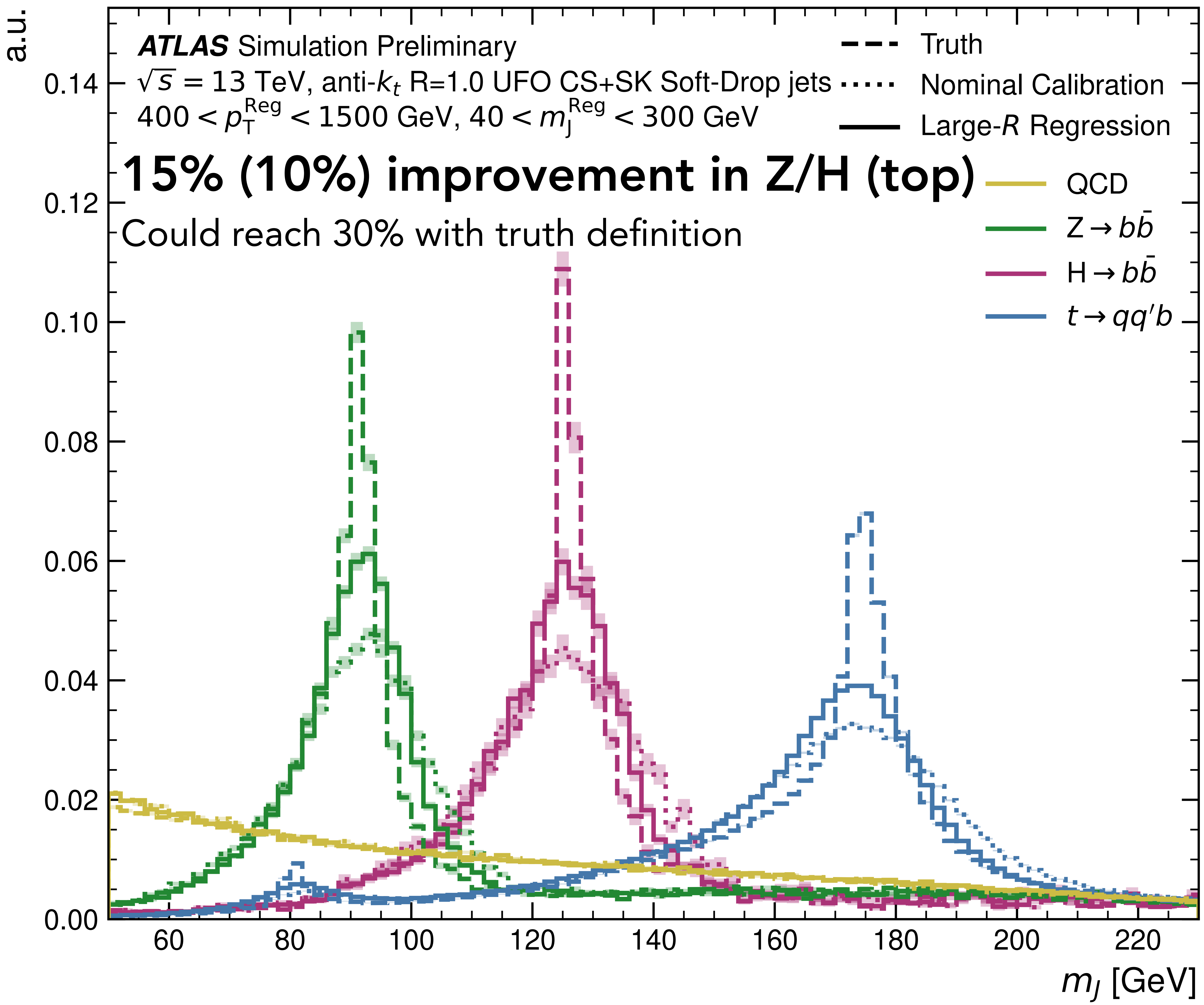
Resonance resolution





- ✦ Evaluation on key di-jet resonances leads to 23% reduction in relative resolution on the Higgs peak!
  - More modest gain of 5% relative to PtReco corrections
  - Can be improved via optimization (e.g. using Z/H samples in training)

Large-R resonance spectrum

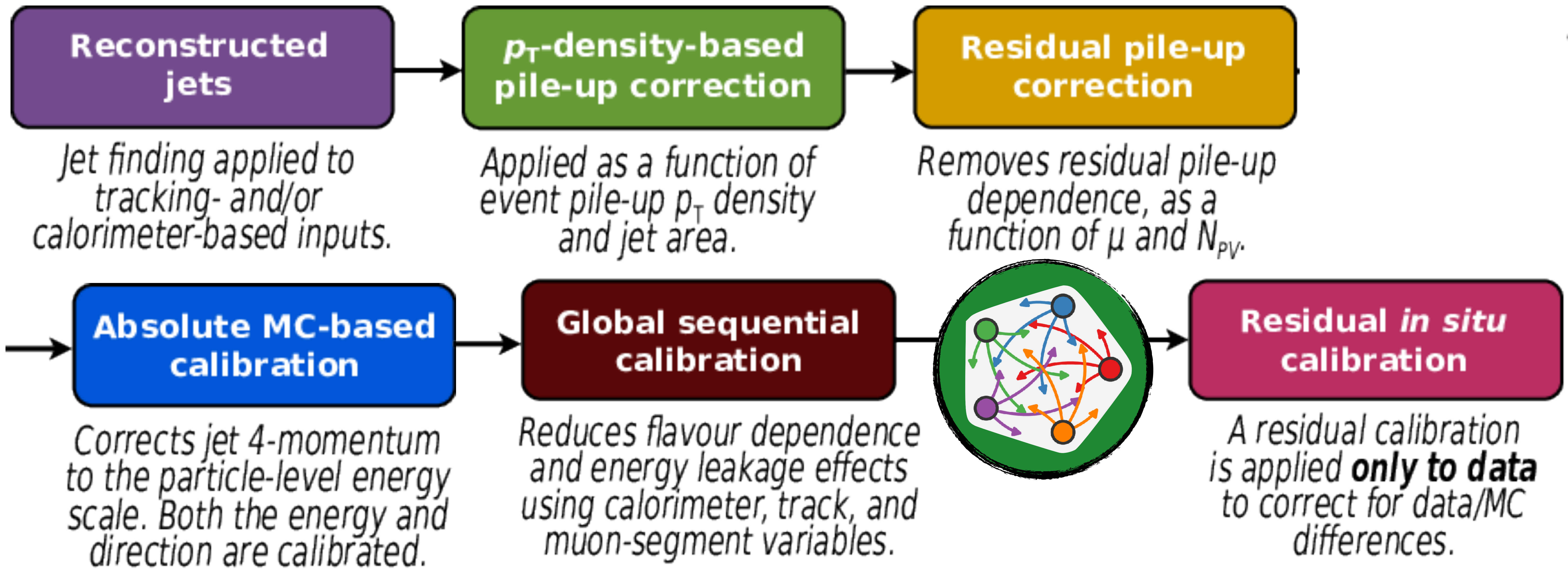


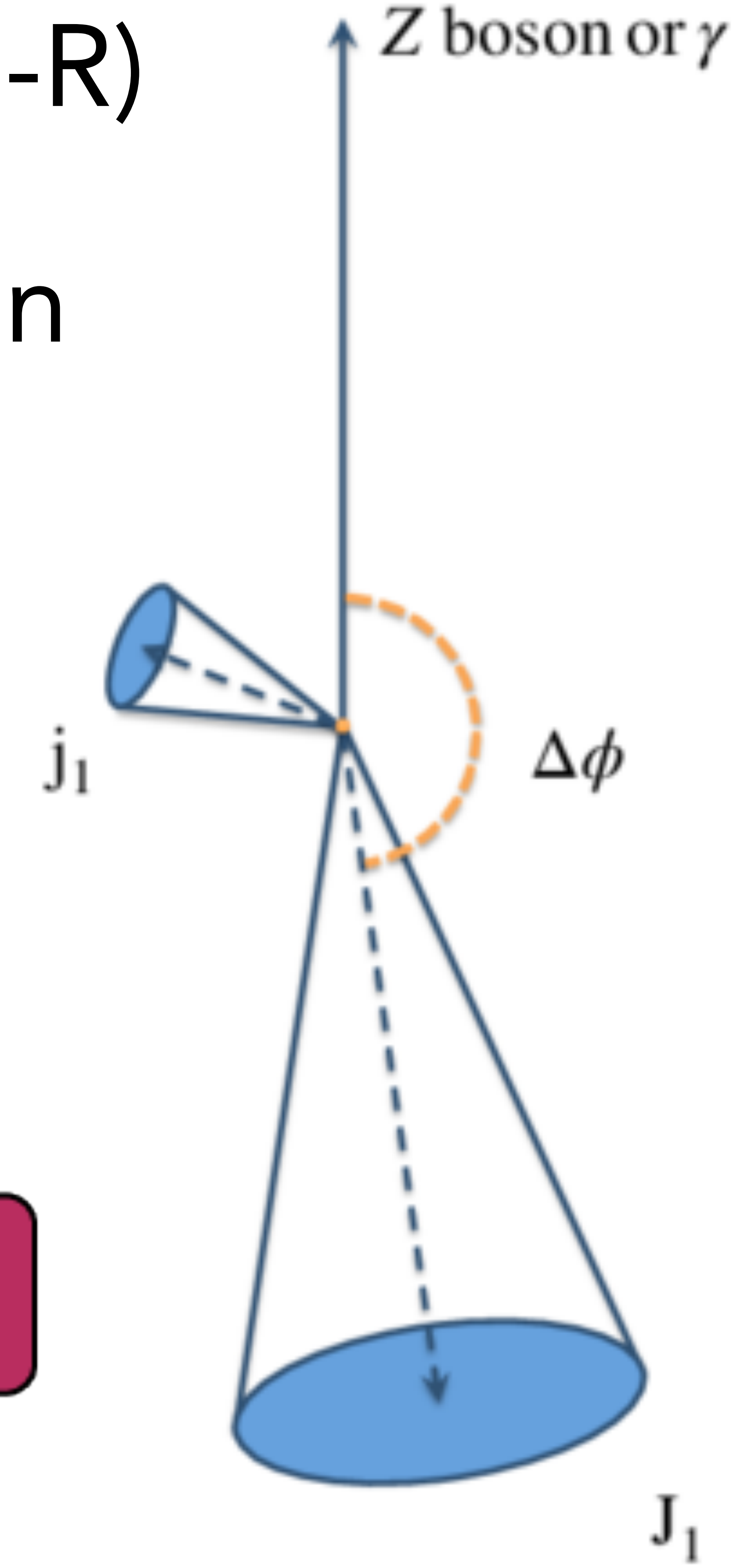
- ✦ Evaluate on SM resonances (Z/H/top)
  - Significant sharpening of Z/H mass peaks
  - Still a long way to go to reach truth-level
- ✦ No mass sculpting in the QCD continuum
  - Use of flat-mass samples eliminates SM mass point bias

Outlook

- Further algorithm optimizations (including lepton corrections for large-R, adding flow objects to small-R)

- So far the model has only been trained on simulation
  - Good performance across processes+algorithms, but **needs calibration** on data for ultimate use for physics!





Thanks for

your attention!

Training datasets

Small-R

Process	Generator	Parton shower	PDF set
Training, validation and test samples			
$pp \to t\bar{t}$ fully/semileptonic	POWHEG [29,30,31]	PYTHIA 8.230 [23] with A14 [32]	NNPDF3.0NLO [24]
$pp \to Z(\mu\mu)$+jets†	MADGRAPH5_AMC@NLO [33] + FxFx [34]	PYTHIA 8.245 with A14	NNPDF3.0NLO
Evaluation samples			
$pp \to Z(\ell\ell)H(b\bar{b})$	POWHEG BOX v2 [29,30,31] + MINLO [35,36,37]	PYTHIA 8.230 with AZNLO [38]	NNPDF3.0NLO
$pp \to Z(\ell\ell)Z(b\bar{b})$	SHERPA [27]	SHERPA 2.2.11	NNPDF3.0NNLO

Large-R

Jet type	Process	Event generator and tune	PDF set
Training, validation and test samples			
$H(b\bar{b})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	PYTHIA 8.306 [23] with A14 [32]	NNPDF3.0NLO [24]
$H(c\bar{c})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	PYTHIA 8.306 with A14	NNPDF3.0NLO
QCD †	Multijet	PYTHIA 8.235 with A14	NNPDF2.3LO
QCD ($b\bar{b}$) ‡	Multijet ($b\bar{b}$), $N_{\text{jet}} \geq 4$, $N_{b-\text{jet}} \geq 2$	PYTHIA 8.235 with A14	NNPDF2.3LO
Evaluation samples			
$H(b\bar{b})$	$q\bar{q}/gg \to ZH$, $Z \to \ell\bar{\ell}/\nu\bar{\nu}/q\bar{q}$	POWHEG v2 +PYTHIA 8.212 [30] with AZNLO [38]	NNPDF3.0NLO
Top	$Z' \to t\bar{t}$	PYTHIA 8.235 with A14	NNPDF2.3LO
$Z(b\bar{b})$	$Z \to b\bar{b}$	SHERPA 2.2.11 [27]	NNPDF3.0NNLO
QCD †	Multijet	PYTHIA 8.235 with A14	NNPDF2.3LO

Model input features

Jet feature	Description
p_{T}	Transverse momentum
η	Signed pseudorapidity
m ‡	Jet mass

Track & charged UFO feature	Description
q/p	Track charge divided by reconstructed momentum
$\mathrm{d}\eta$	Pseudorapidity of track relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Transverse IP: Closest distance from track to beam-line in the transverse plane
$z_0 \sin\theta$	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Significance of transverse IP
$s(z_0 \sin\theta)$	Significance of longitudinal IP times the sin of the polar angle
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
LeptonID †	Information on if the track was used in lepton reconstruction

Charged & neutral UFO feature	Description
$p_{\mathrm{T}}^{\mathrm{Flow}}$ ‡	Transverse momentum of charged flow constituent
E_{Flow} ‡	Energy of charged flow constituent
$\mathrm{d}\eta_{\mathrm{Flow}}$ ‡	Pseudorapidity of track relative to the large-R jet η
$\mathrm{d}\phi_{\mathrm{Flow}}$ ‡	Azimuthal angle of the track, relative to the large-R jet ϕ
$\mathrm{d}r_{\mathrm{Flow}}$ ‡	Angular distance of the track from the large-R jet direction

Soft Muon Input	Description
p_{T}	Transverse momentum
η	Signed pseudorapidity
ϕ	Azimuthal angle
dR	Angular distance of the soft muon from the small-R jet axis
q/p	Muon charge divided by the reconstructed momentum
Momentum Balance Significance	Ratio of the difference in momentum measured by the ID and MS to the uncertainty on the energy loss measured by the calorimeters
Scattering Neighbour Significance	Sum of the significances of the angular difference $\Delta\phi$ between pairs of adjacent hits along the track, multiplied by the particle charge
$p_{\mathrm{T}}^{\mathrm{rel}}$	Orthogonal projection of the muon p_{T} onto the jet axis
d_0	Transverse IP: Closest distance from track to beam-line in the transverse plane
z_0	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$\sigma(d_0)$	Uncertainty on measurement of transverse IP
$\sigma(z_0)$	Uncertainty on measurement of longitudinal IP
$d_0/\sigma(d_0)$	Significance of transverse IP
$z_0/\sigma(z_0)$	Significance of longitudinal IP

Soft Electron Input	Description
$p_{\mathrm{T}}^{\mathrm{r}}$	Relative p_{T} of the electron with respect to the jet
dR	Angular separation between electron and jet axis
$p_{\mathrm{T}}^{\mathrm{iso}}$	Isolation variable
$\|\eta\|$	Absolute value of pseudorapidity
$s(d_0)$	Transverse IP: Closest distance from track to beam-line in the transverse plane
$z(d_0)$	Longitudinal IP: Closest distance from track to PV in the longitudinal plane
$s(d_0/\sigma_{d_0})$	Significance of the transverse IP
$\Delta\phi^{\mathrm{res}}$	The azimuthal angle difference $\Delta\phi$ between the cluster position in the middle layer and the track.
E/p	Ratio of the cluster energy to the track momentum
R_{had}	Ratio of E_{T} in the hadronic calorimeter to E_{T} of the EM cluster
$R_{\mathrm{had\,1}}$	Ratio of transverse energy E_{T} in the first layer of the hadronic calorimeter to E_{T} of the EM cluster
E_{ratio}	Ratio of the energy difference between the largest and second-largest energy deposits in the cluster over the sum of these energies
$w_{\eta 2}$	Lateral shower width
R_{η}	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position
f_1	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter
f_3	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter
p_{HF}	Probability of being from heavy flavour decay