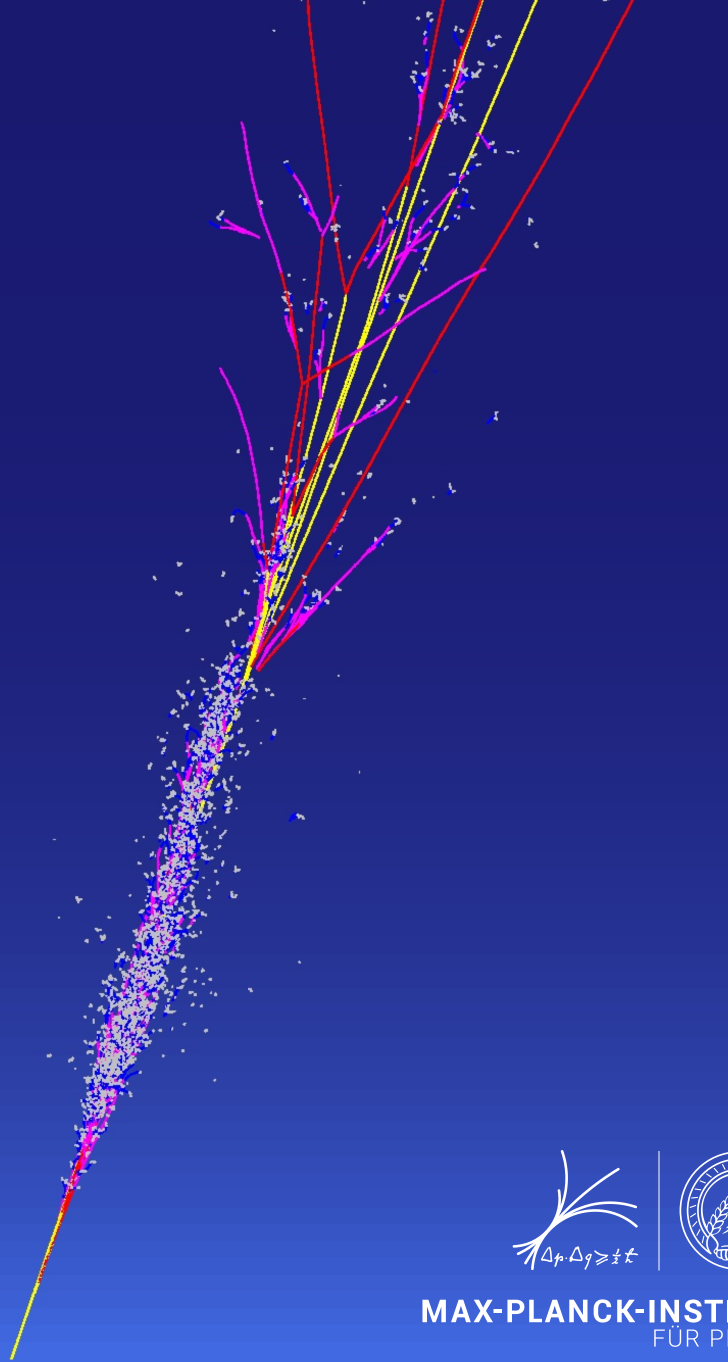
New techniques for reconstructing, calibrating and identifying hadronic objects with ATLAS

Blois2024 | Sven Menke, MPP München | 24 Oct 2024, Blois

- Hadronic objects in ATLAS
  - Importance of Pile-Up
  - Topological clustering
- Time as a new discriminant
- Calibration methods
  - ML for cluster calibration
  - ML for jet energy calibration
  - ML for energy and mass for large $\triangle R$-jets
- Identification methods
  - Quark/gluon tagger
  - W-boson tagger
  - Top-quark tagger
- Performance of missing transverse momentum
- Conclusions



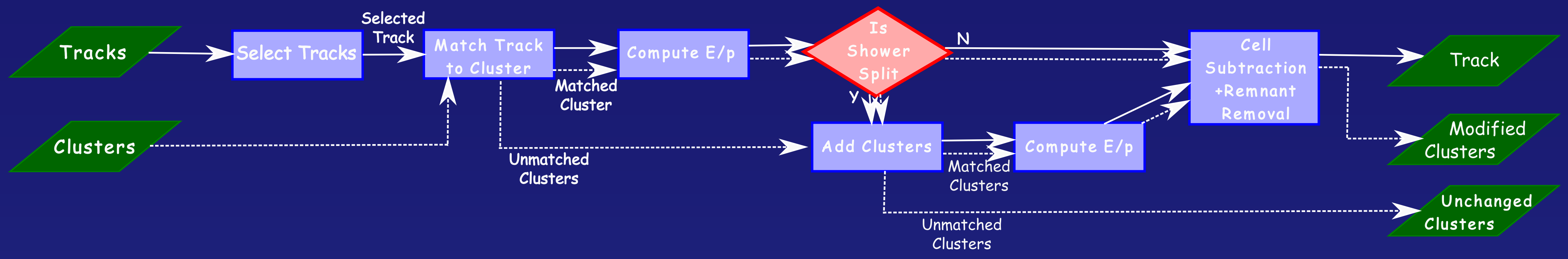
ATLAS EXPERIMENT

MAX-PLANCK-INSTITUT FÜR PHYSIK

Hadronic objects in ATLAS

- Jets in ATLAS are made out of topological clusters (calorimeter) and charged particle tracks (inner detector)
- Clusters and tracks are combined to form higher level objects (with 4-vectors) as input to jet-clustering
  - Particle Flow Objects (PFO) for small and large ΔR jets (see sketch)

Eur.Phys.J.C77(2017)466

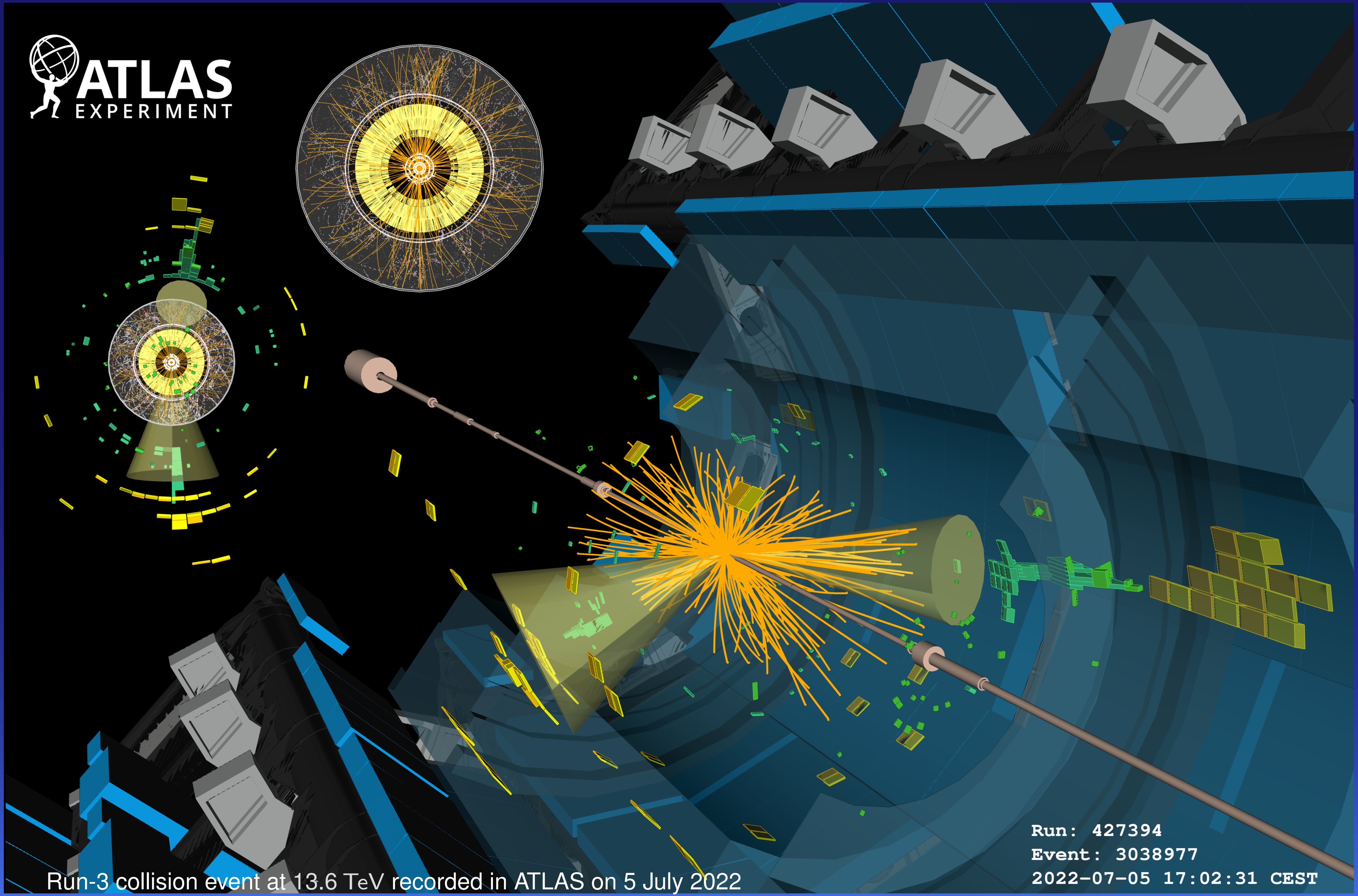


Flow chart particle flow algorithm

  - Track-CaloClusters (TCC) for large ΔR jets (splitting the cluster energy to all matching tracks with track's $p_\perp$-fraction in matched tracks as weight, ATL-PHYS-PUB-2017-015)
  - Unified Flow Objects (UFO) for large ΔR jets combine PFO and TCC depending on environment to make best of both Eur.Phys.J.C81(2020)334)
- Jet clustering is performed with FastJet – anti-k_t with $\Delta R = 0.4$ (small) or $\Delta R = 1.0$ (large)
- Jets are then calibrated in several steps for energy ($p_\perp$), momentum direction and mass (for large ΔR)

Di-jet event in Run-3

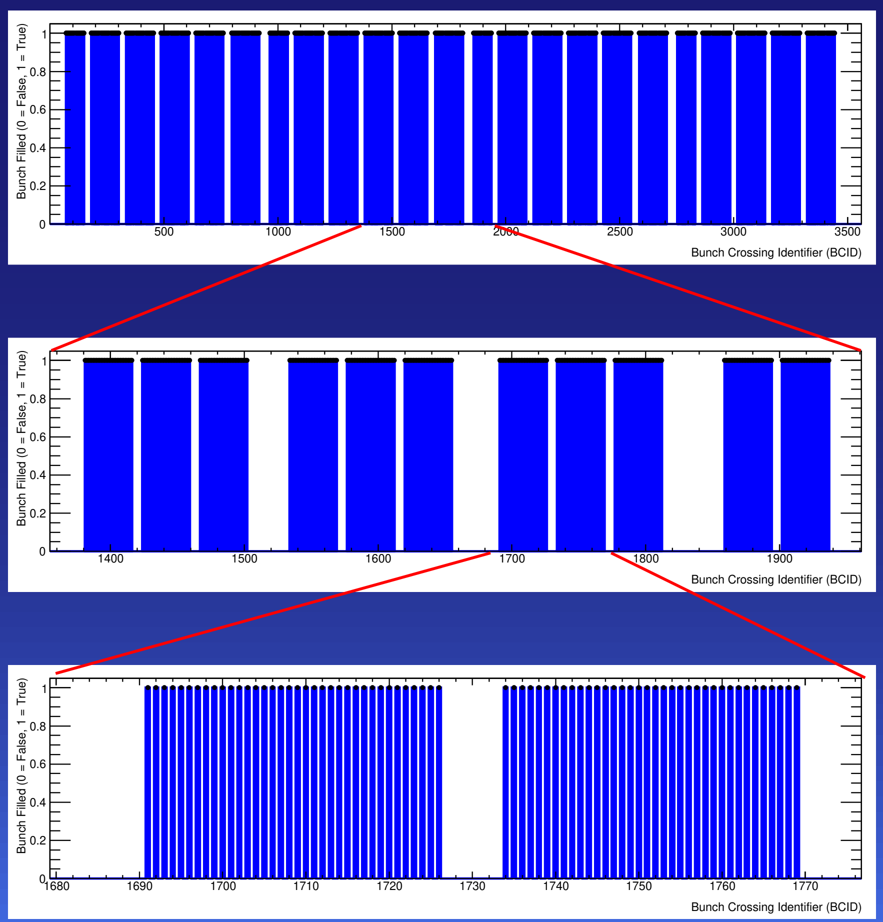
twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun3Collisions



Pile-Up impact on calorimeter signals in ATLAS

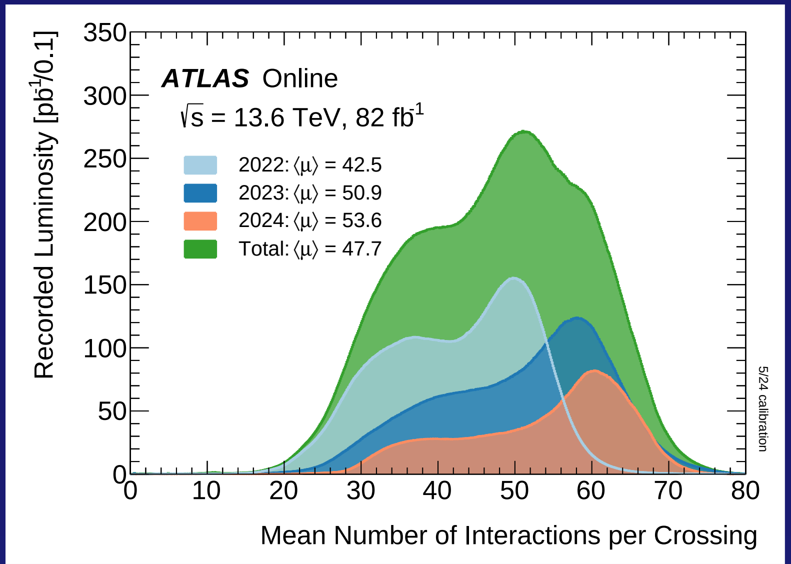
- Pile-Up characteristics
  - μ: Average number of interactions per crossing $\sim$ 50 in Run-3
  - Δt: Bunch distance 24.95 ns
  - Signal integration time for the LAr-calorimeters $\sim$ 500 ns

26.7 km $\equiv$ 3564 bunch places each Δt = 24.95 ns



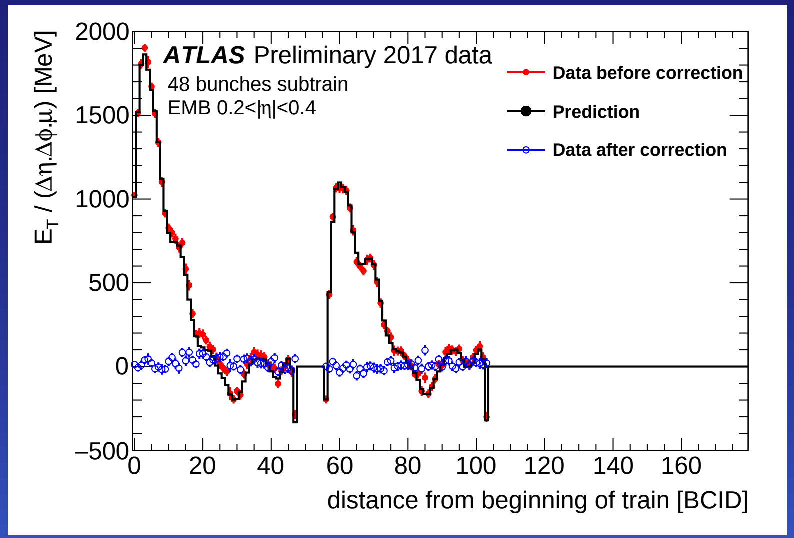
- Typical LHC bunch structure in 2022-2024
  - 2340 colliding bunch pairs (2835 is theoretical max)
  - 23 trains $\sim$ 800 ns apart (LHC injection)
  - 2-3 sub-trains with gaps of 200 ns (SPS injection)
  - 36 filled bunches per sub-train
    - Pile-Up impact depends on bunch crossing Number (BCID)
    - Up to 20 colliding bunch pairs contribute to signal
- See **arXiv:2407.10819**, where ATLAS turns noise into data: Using pileup for physics

twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3



Interactions per crossing 2022-2024

twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResults2015

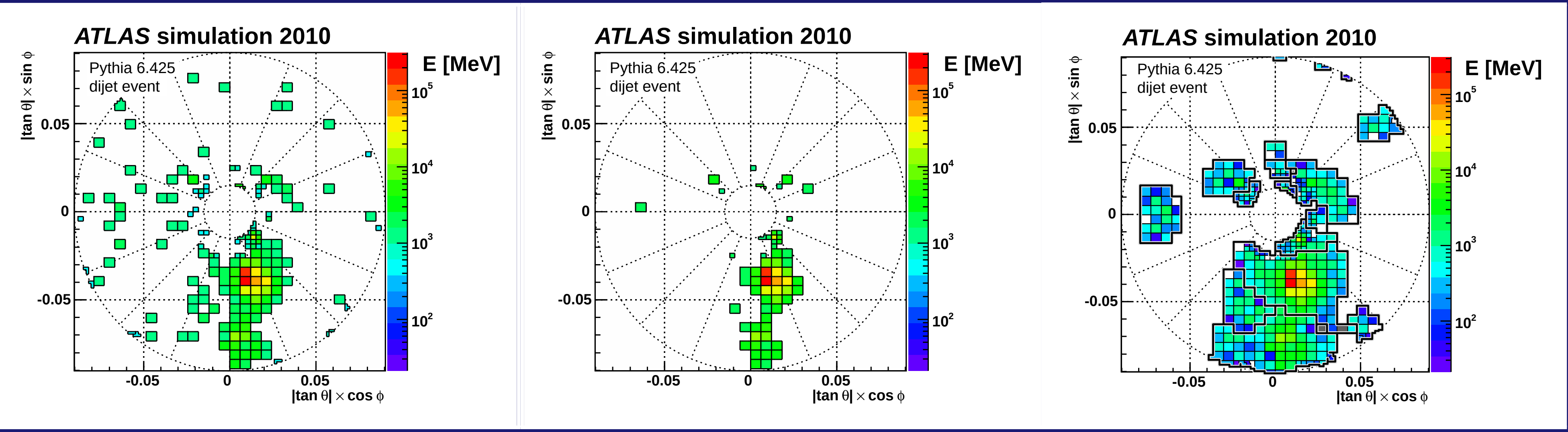


LAr baseline shift

Topological clustering

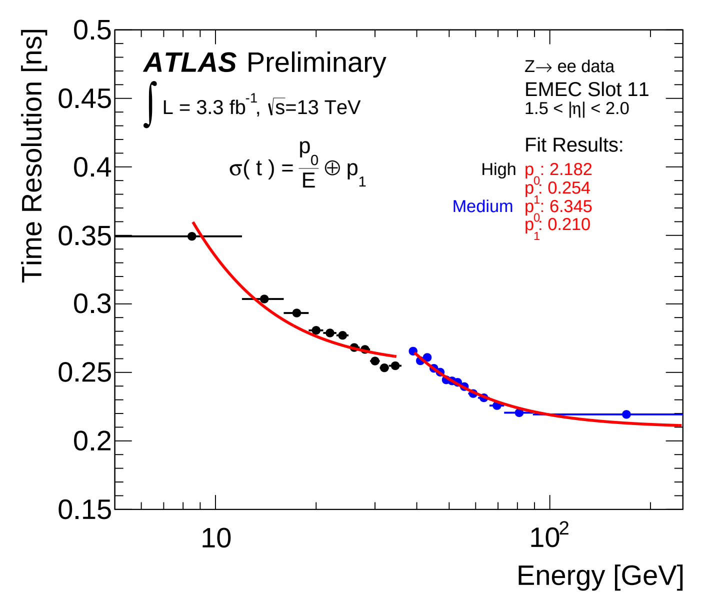
- Jet constituents, $\tau^{\pm}$, $e^{\pm}$ and γ are made out of topological cell clusters (TopoClusters)
  - 3d energy blobs of neighbouring calorimeter cells around seeds with $|E| > 4\sigma$
  - Direct seed neighbours with $|E| > 2\sigma$ become seeds too
  - Proto-clusters are re-clustered around local energy maxima
  - σ is the expected noise $\equiv \sigma_{\text{elec}} \oplus \sigma_{\text{pile-up}}$

Eur.Phys.J.C77(2017)490



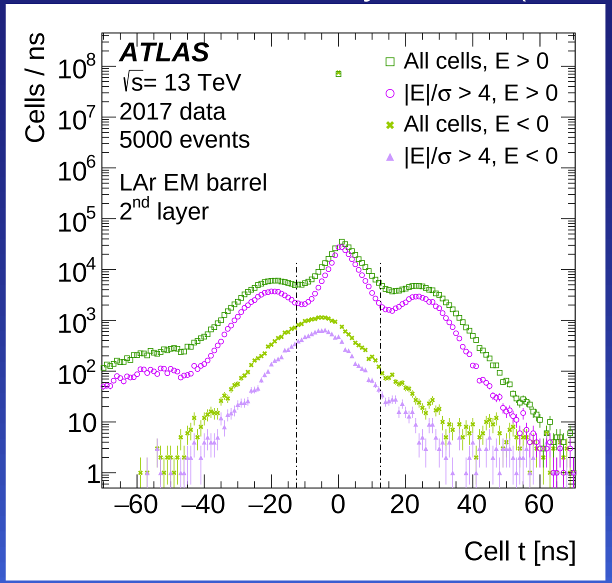
$|E| > 2\,\sigma_{\text{noise}}$ $|E| > 4\,\sigma_{\text{noise}}$ 4 / 2 / 0 TopoClusters

twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResults2015



LAr EMEC time resolution

Eur.Phys.J.C84(2024)455

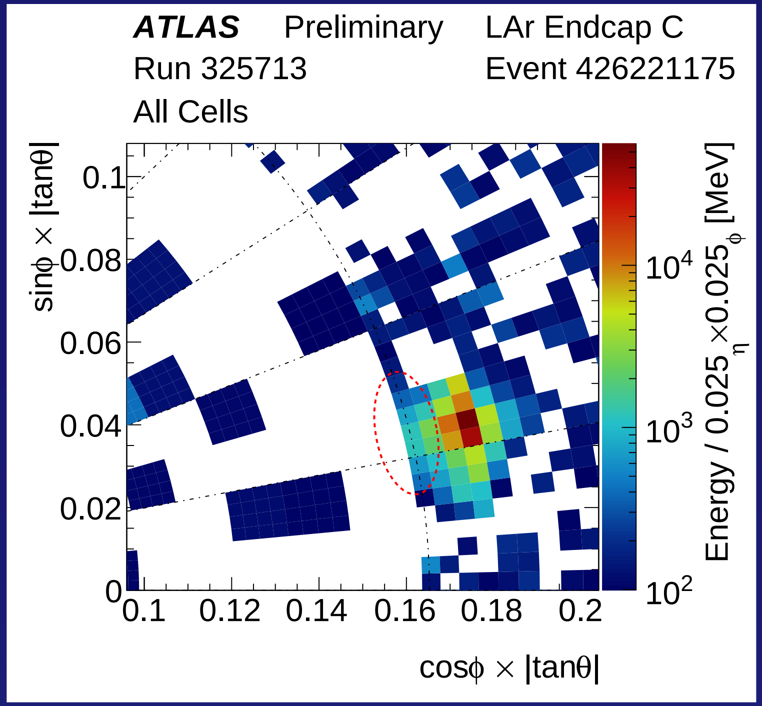


Cell time in LAr EMB

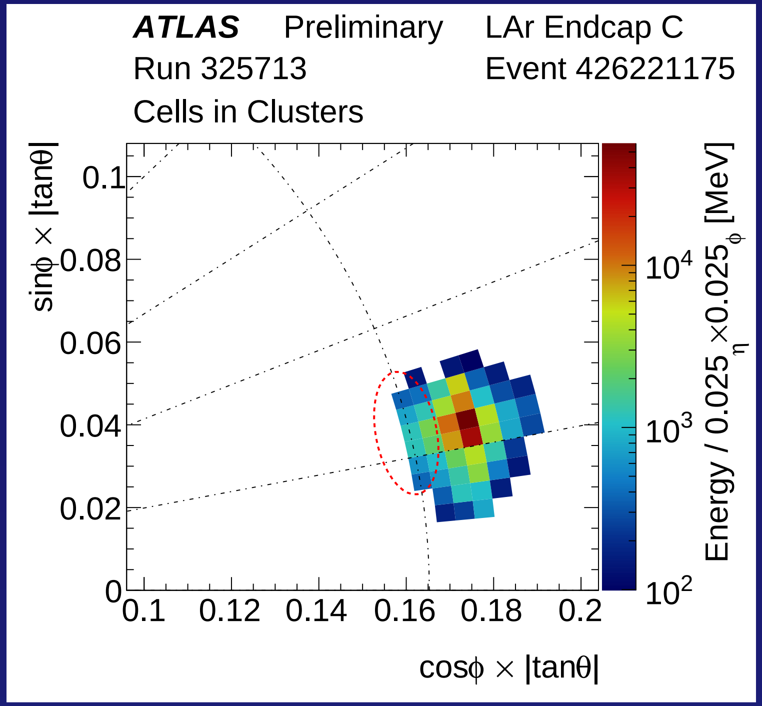
- Calorimeters have excellent time resolution!
  - Intrinsic time resolution in LAr samplings is $\sim$ 60 ps at high energies
  - Time has always been reconstructed alongside energy since the beginning of data taking
  - Added recently to the topological clustering algorithm as additional discriminator (cut at $|t| < 12.5$ ns) for any cell that has $|E| > 4\sigma$
  - But restrict the time cut to those cells with $E < 20\sigma$
    - To keep significant, positive energy deposits that are out-of-time (searches for exotic, long-lived particles)

Time as a new discriminant

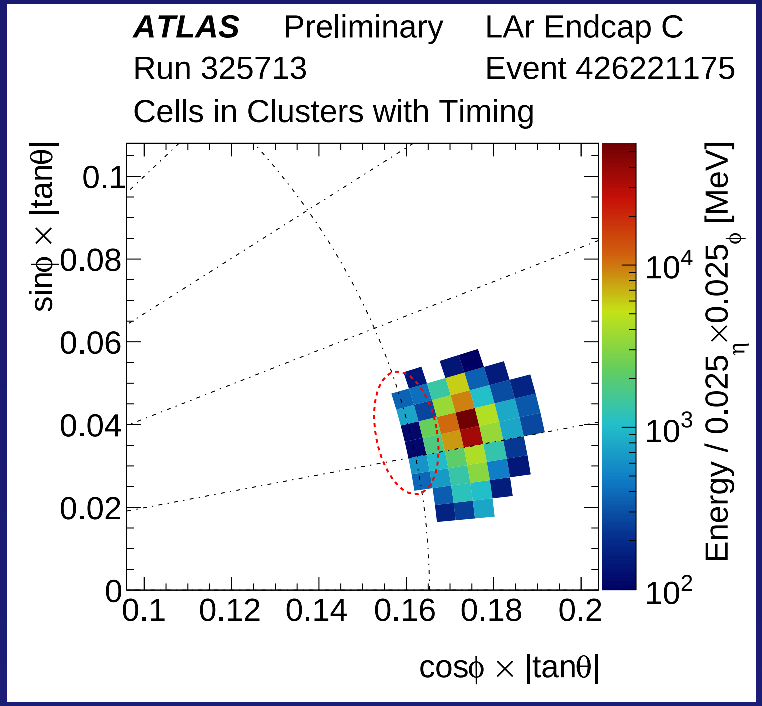
twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResults2015



LAr cell energy sums above 100 MeV



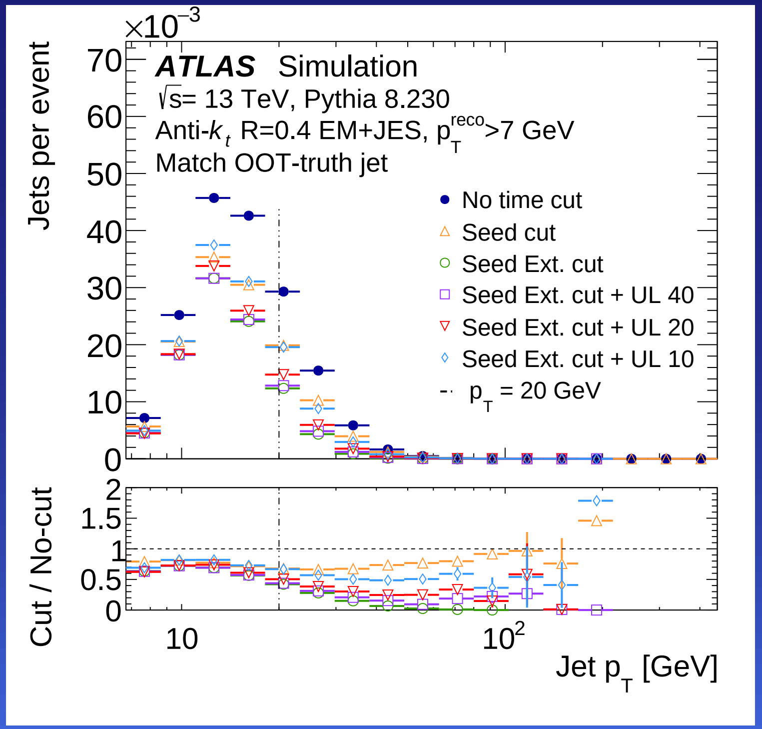
Inside default clusters



Inside clusters with time discrimination

- New default in Run-3

Eur.Phys.J.C84(2024)455



OOT Pile-Up jets vs. $p_\perp$

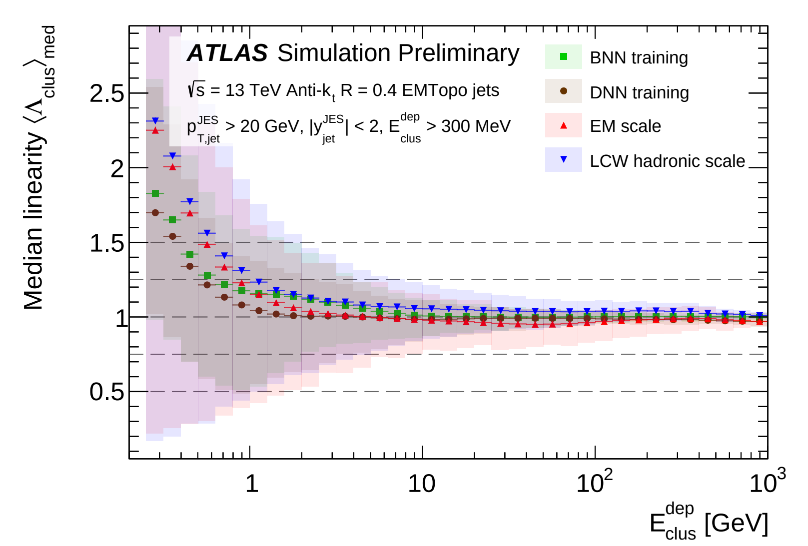
- New time discriminant further reduces residual out-of-time (OOT) Pile-Up that was not suppressed by default topological clustering
  - Entire clusters are removed (reduces background)
  - And cells inside clusters are removed (improves signal, see plots above)
- removes OOT Pile-Up jets (see plot to the right)
  - By $\sim 50\%$ at $p_\perp \simeq 20$ GeV; by $\sim 80\%$ at $p_\perp \geq 50$ GeV
  - Number of in-time jets remains unchanged
  - Resolution improves by $\sim 5\%$
- Removes fakes for $\tau^\pm$, $e^\pm$ / γ

Calibration methods

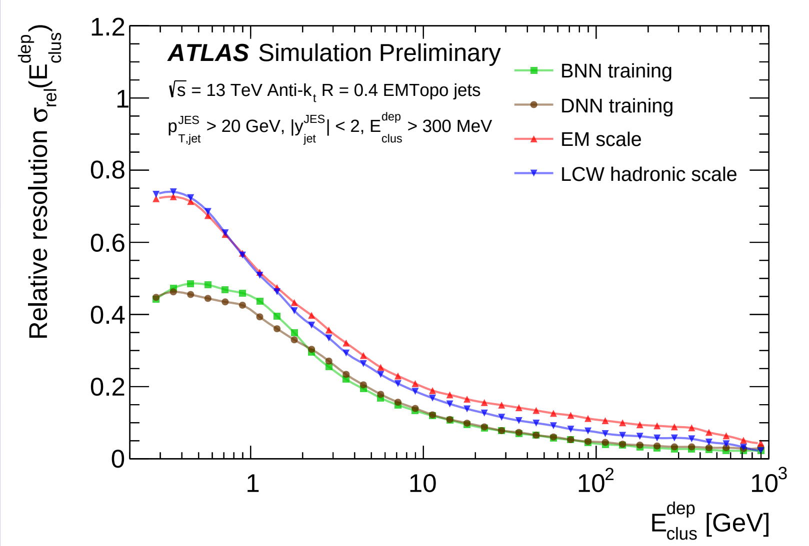
- Local Hadronic Calibration (a.k.a. Local Cell Weighting, LCW, Eur.Phys.J.C77(2017)490)
  - 4 step procedure to bring the energy scale of clusters from the raw "EM"-scale to the particle-level "LC"-scale
  - Classification: Compute EM-probability p^{EM} form shapes
  - Cell-weighting: Apply hadronic (HAD) and electromagnetic (EM) weights:
    - $w_{\mathrm{cell}} = (1 - p^{\mathrm{EM}})\,w_{\mathrm{HAD}} + p^{\mathrm{EM}}\,w_{\mathrm{EM}}$
  - For 3 different corrections:
    - Corrections for hadronic non-compensation
    - Corrections for out-of-cluster deposits
    - Corrections for out-of-calorimeter (dead-material) deposits
- Jets (Eur.Phys.J.C81(2021)689)
  - Can use either EM- or LC-scale objects (clusters or flow objects)
  - Are corrected for Pile-Up (jet-area correction and residual Pile-Up correction)
  - Get their energy corrected by MC-derived Jet-Energy-Scale correction
  - Flavour dependency and resolution gets improved by Global-Calibration (MC-derived, keeping average energy scale constant)
  - Data is corrected *in-situ* from measured $p_{\perp}$ balance of jets in multi-jet and Z^0/γ + jet events to match MC

Cluster calibration with neural networks

- Idea: Apply machine learning to Local Hadronic Calibration
  - To explore the applicability of neural networks to calorimetric calibration
  - So far done with the first of the three correction steps (non-compensation) and implicit classification
    - Biggest difference to legacy LCW: Pile-Up is included
  - Out-of-cluster and dead-material corrections still to come
- Input quantities for the NNs:
  - Kinematics: $(E^{\mathrm{EM}}_{\mathrm{clus}}, y^{\mathrm{EM}}_{\mathrm{clus}})$ • Significance: $(E_{\mathrm{clus}} / \sigma_{\mathrm{clus}})$ • Time: $(t_{\mathrm{clus}}, \mathrm{Var}_{\mathrm{clus}}(t_{\mathrm{cell}}))$ • Cluster moments: (depth, centroid, EM-fraction, energy density, lateral and longitudinal dispersion, compactness) • Environment: (isolation, N_{PV}, μ)



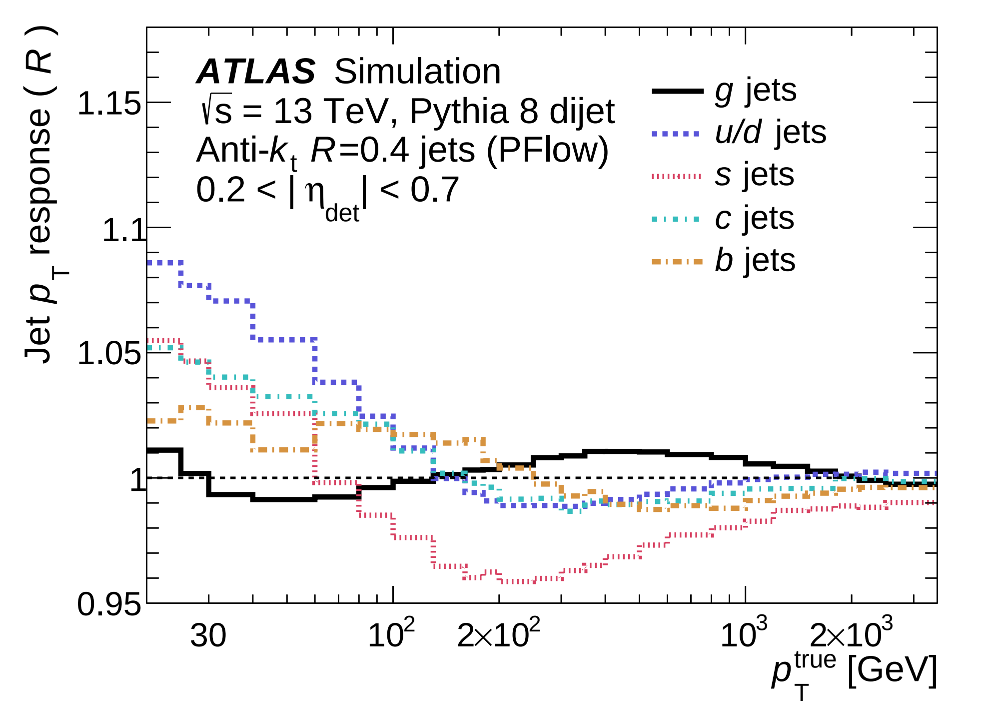
Median linearity of reconstructed $E^{\mathrm{dep}}_{\mathrm{clus}}$



Relative resolution of reconstructed $E^{\mathrm{dep}}_{\mathrm{clus}}$

- Trained NNs:
  - Deep Neural Net (DNN) with leaky Gaussian kernel
  - Bayesian Neural Net (BNN) with regularised negative log-likelihood
- Linearity (left) and resolution (right) of NNs compared to EM- and LC-scale on simulated clusters from di-jets with Pile-Up
  - NNs outperform legacy LCW (removal of Pile-Up)
    - but Pile-Up removal is not part of LCW ...
  - DNN slightly better than BNN
  - Encouraging result to implement the other steps

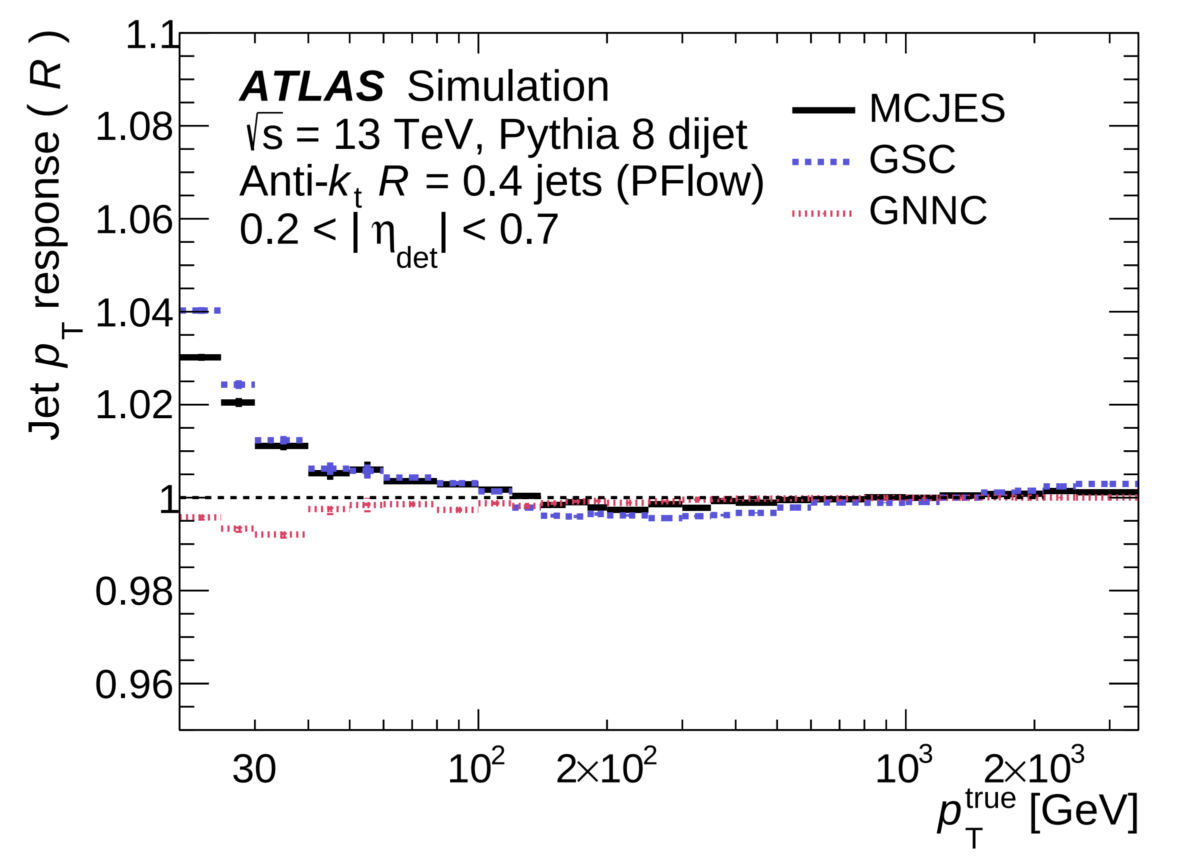
 Eur.Phys.J.C83(2023)761



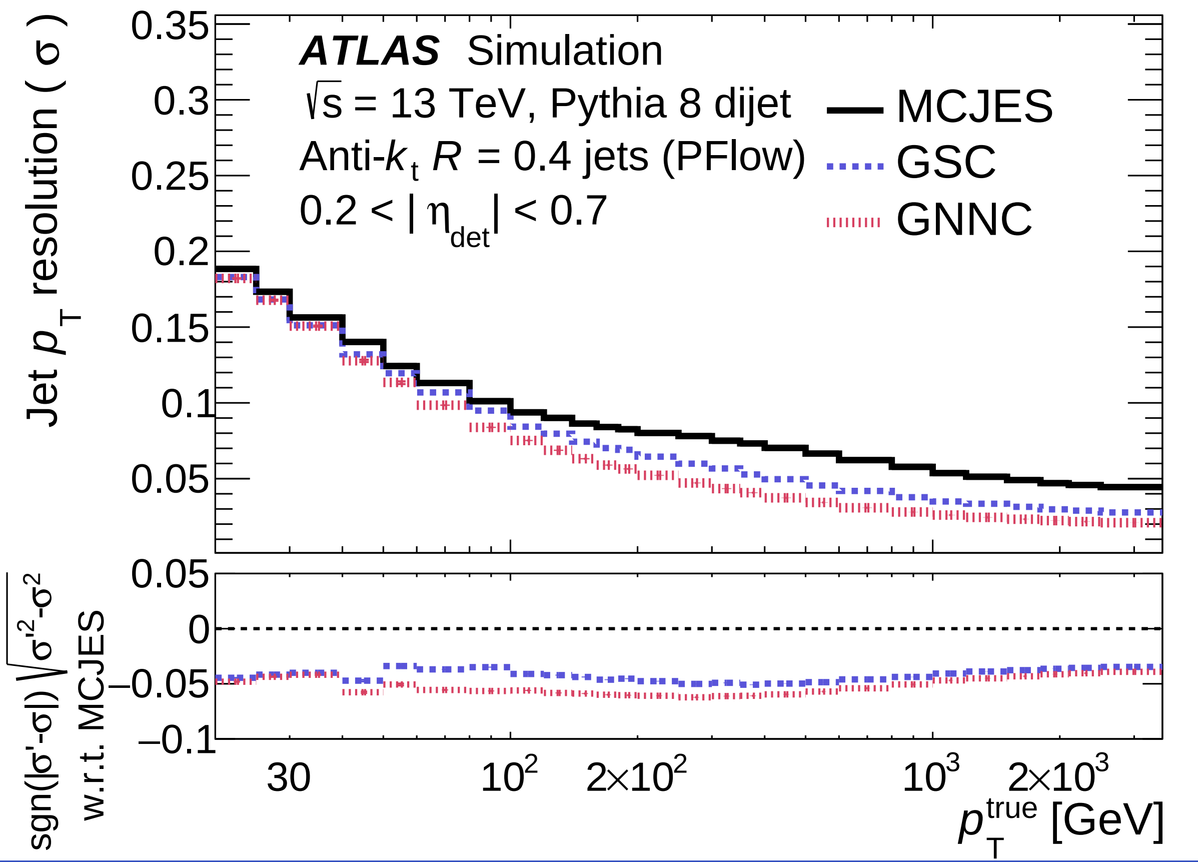
$p_\perp$ response after MCJES

- Global-Calibration is applied to jets after setting the jet energy scale (MCJES) (based on MC simulations and energy E and pseudo-rapidity η of the jet)
- Global Sequential Calibration (GSC) (used for Run-2)
  - Uses many kinematic observables in addition to $p_\perp$: • Charged $p_\perp$ fraction f_{charged} • Energy fractions in first Tile & third EM layer f_{Tile0} • f_{LAr3} • Number of tracks N_{track} • $p_\perp$-weighted average track distance w_{track} • Number of associated muon segments N_{segments}
  - Since JES is kept unchanged, the six corrections can be applied and checked independently of each other ► Requires uncorrelated observables
- Global Neural Network Calibration (GNNC) (new, will be used for Run-3)
  - Alternative to GSC
  - Trains a DNN with jet observables for a simultaneous correction to $p_\perp$ and leaky Gaussian kernel loss-function
    ► Allows the use of correlated variables; is allowed to change JES
  - In addition to the GSC observables it uses: • 12 more (i.e. all 14) layer energy fractions $f_{\text{LAr0-3,Tile0-2,HEC0-3,FCal0-2}}$ • Number of clusters with 90% energy $N_{90\%}$ • η • Pile-Up variables μ, N_{PV}

- Closure and resolution in $p_\perp$ compared after MCJES, MCJES+GSC and MCJES+GNNC (here for $0.2 < |\eta| < 0.7$, similar results in all other η-regions)
- Small non-closure for GSC at low $p_\perp$ stems from MCJES (GSC keeps JES unchanged)
- GNNC does change JES and hence improves the MCJES closure at low $p_\perp$
- Resolution improves by 15 - 25% for GNNC compared to GSC

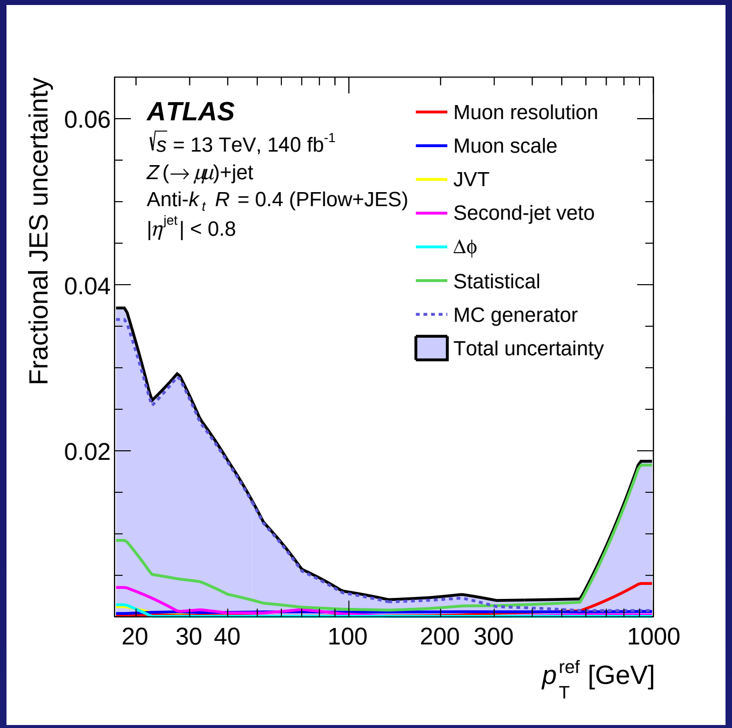


Jet $p_\perp$ response closure

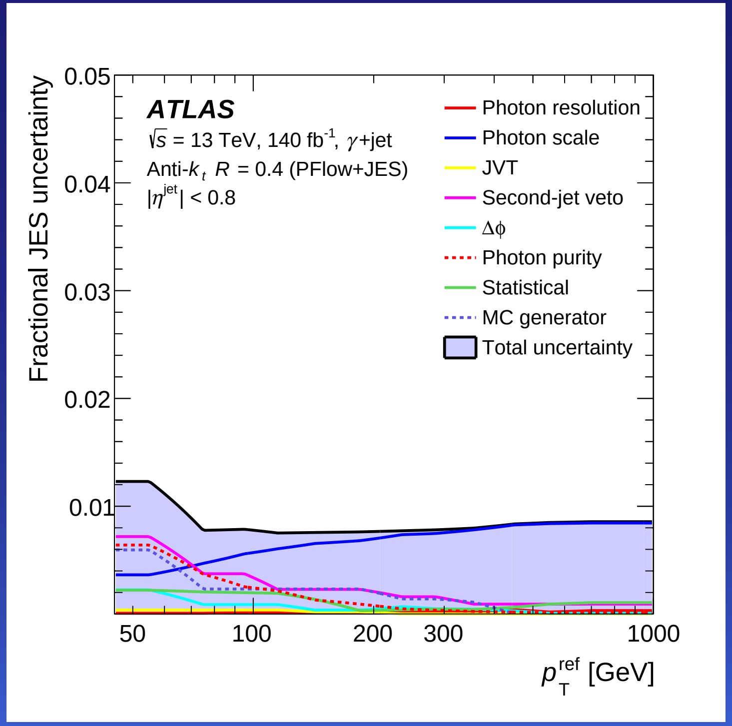


Jet $p_\perp$ resolution

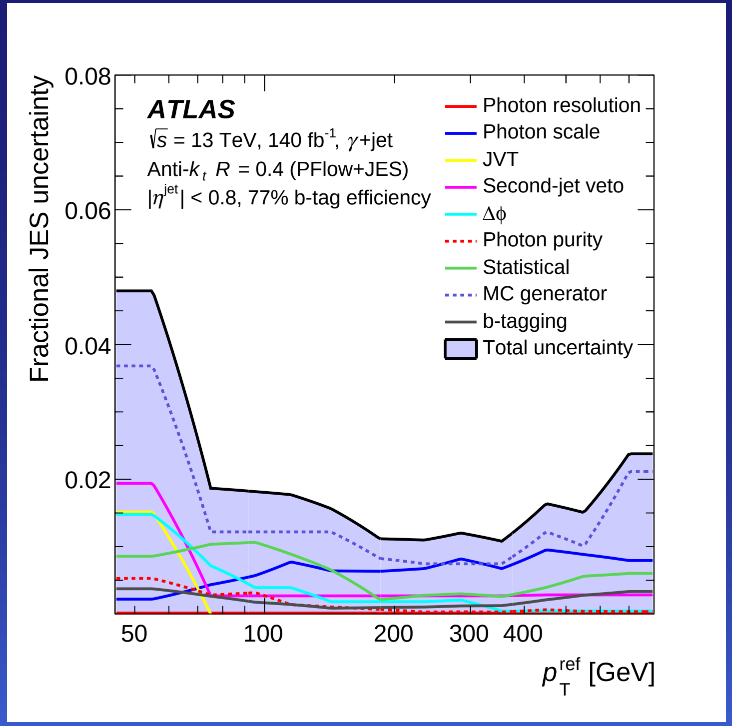
- Z^0 + jet and γ + jet data are compared *in-situ* to simulations to bring the final JES in data to simulation level (after MCJES+GNNC and *in-situ* η-intercalibration with multi-jet events)
- Missing-$E_\perp$ Projection Fraction (MPF) is used to calculate $p_\perp$-balance between Z^0 / γ and the full hadronic recoil
  - best for Pile-Up and lower $p_\perp$
- $O(1\%)$ precision is achieved over a large $p_\perp$-range



JES uncertainty vs. $p_\perp$: MPF method $Z^0(\to \mu^+\mu^-)$ + jet



JES uncertainty vs. $p_\perp$: DB method γ + jet

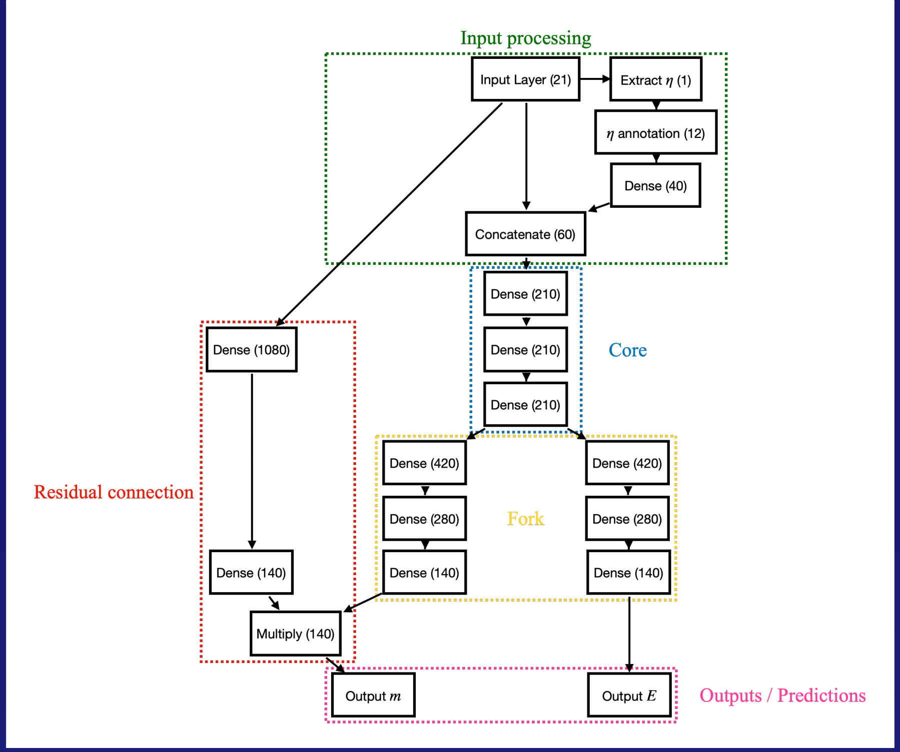


b-jet JES uncertainty vs. $p_\perp$: DB method γ + b-jet

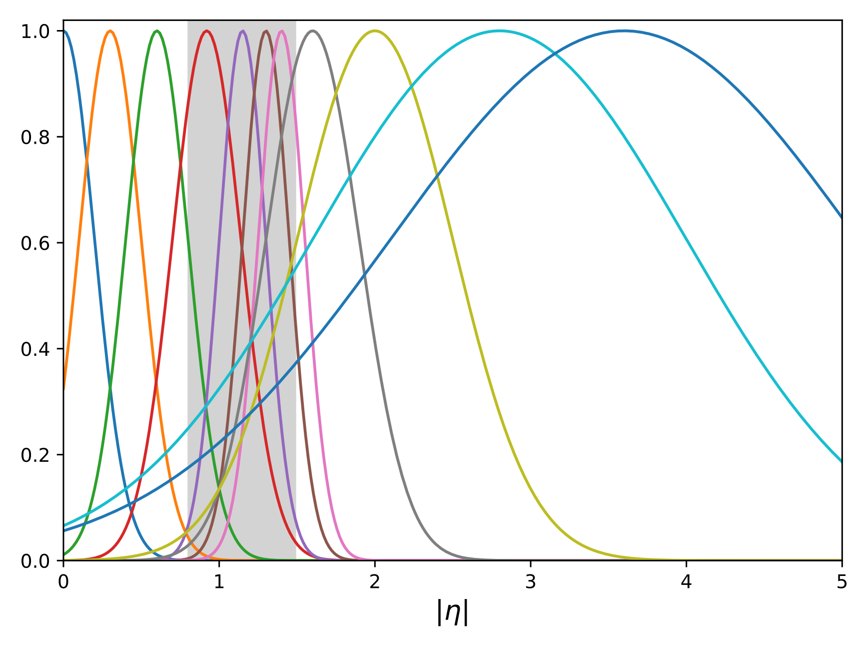
- Direct Balance (DB) is used in γ + jet events to measure the balance of γ against one (possibly b-tagged) jet
  - Good at $p_\perp > 100$ GeV and for single *b*-jets
- Up to $O(1\%)$ precision on-top of general JES uncertainty

Calibration of E and m of large ΔR-jets

- Large ΔR jets: good for boosted topologies of heavy resonances
  - The asymmetric response in energy and mass requires dedicated calibration for both
  - Both remain highly correlated though and a combined calibration approach is hence desirable
- Complex DNN with η annotation (adding 11 Gaussian η-dependent weights to input)
  - Inputs: Jet kinematics E, m, η, 8 jet substructure variables, 7 detector-level energy or $p_\perp$ fractions, Pile-Up environment N_{PV}, μ
  - Initial training for both E and m
  - Loss function is sum of negative log-likelihood predicting μ and σ of Gaussian distributions in E and m
  - Then fork and optimise separately for E and m (can freeze the other)
  - Residual connection for m improves the focus on most important inputs for m
- Trained on 270 M jets from fully simulated di-jet events (based on Pythia8 and Geant4; other generators, physics for cross-checks)

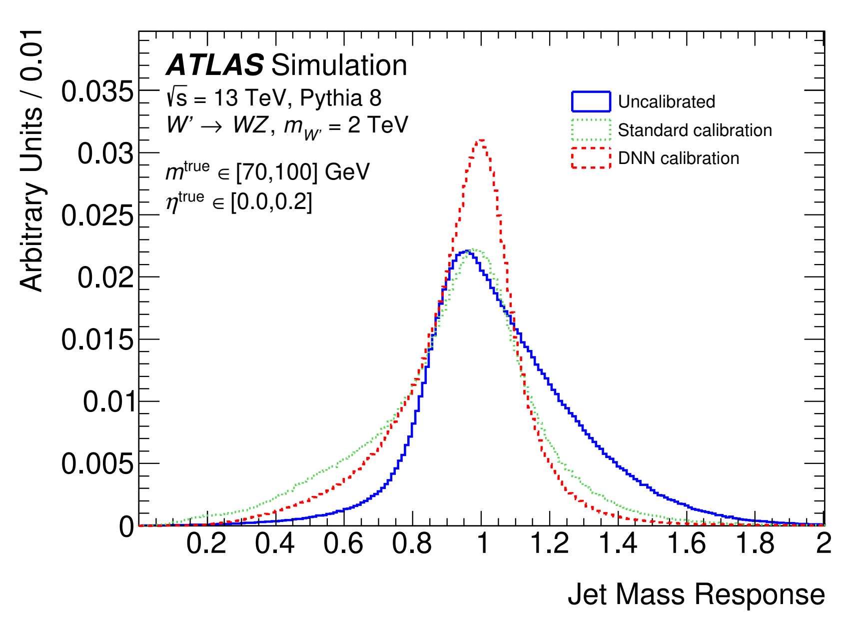


DNN architecture for E and m

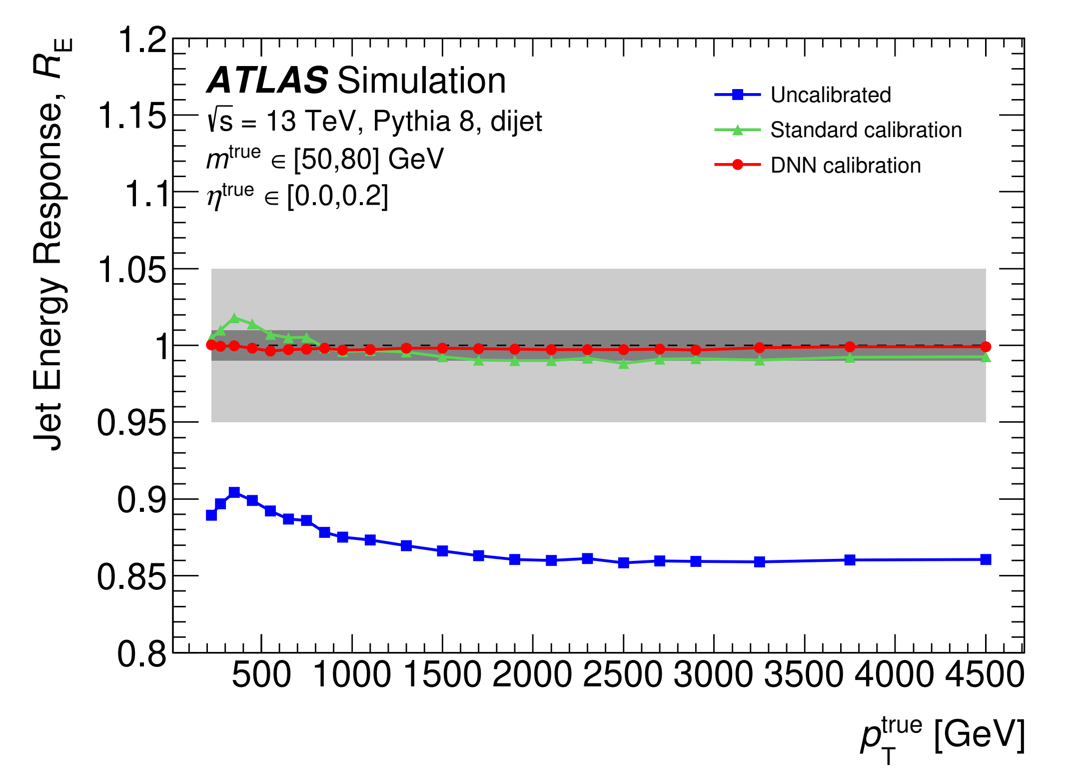


η-annotation functions

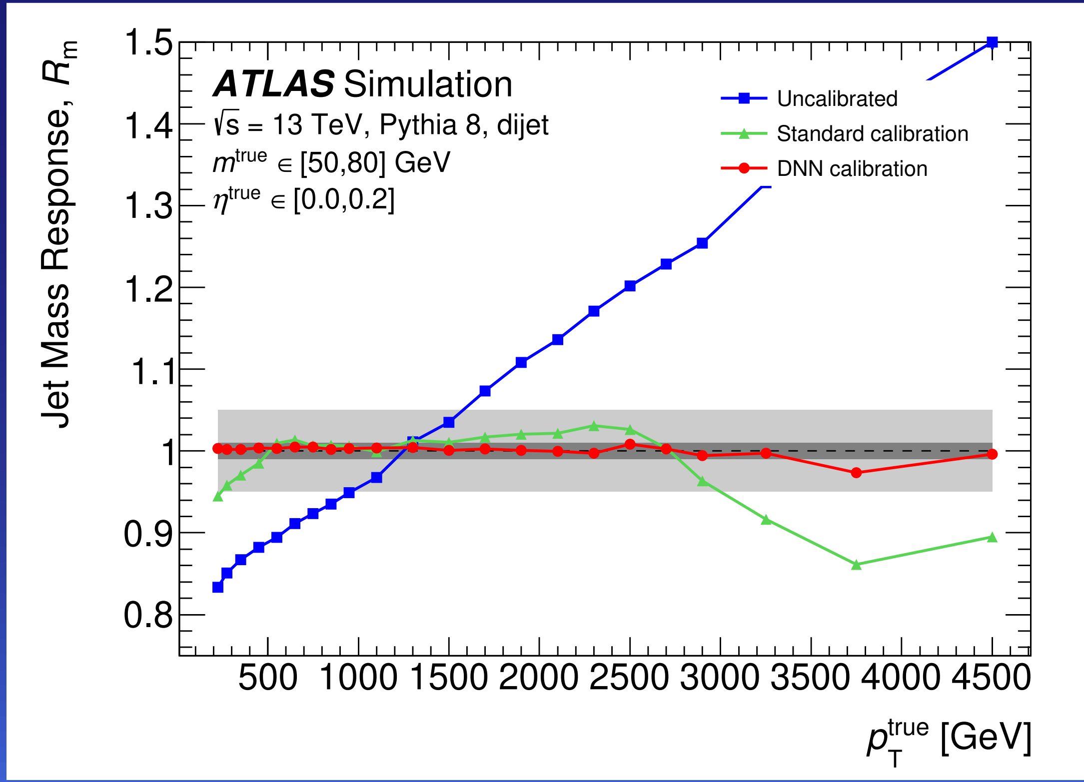
- Comparison of the DNN calibration (red) with standard calibration (green) and no calibration (blue)
  - DNN outperforms standard calibration in energy- and mass-scale closure and resolution for both E and m
  - Typical resolution improvement of $> 30\%$ for $p_{\perp} > 500\,\mathrm{GeV}$
  - Robust against Pile-Up
  - Performs also better on topologies not used in the training (boosted heavy bosons)



Mass response for boosted $W^{\pm}$ / Z^0



Energy response for different calibrations

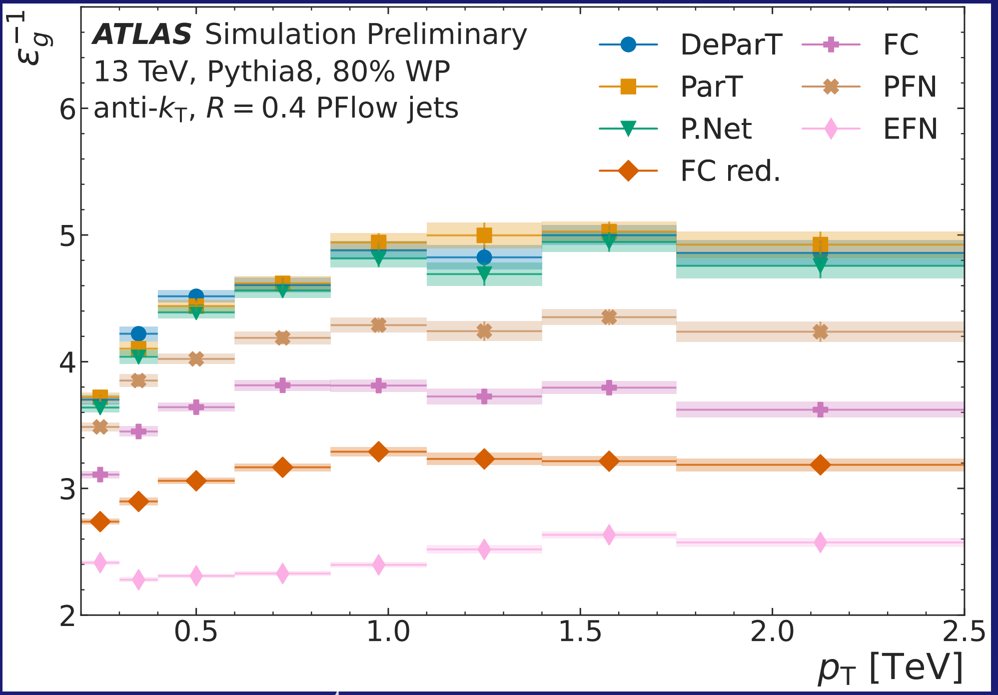


Mass response for different calibrations

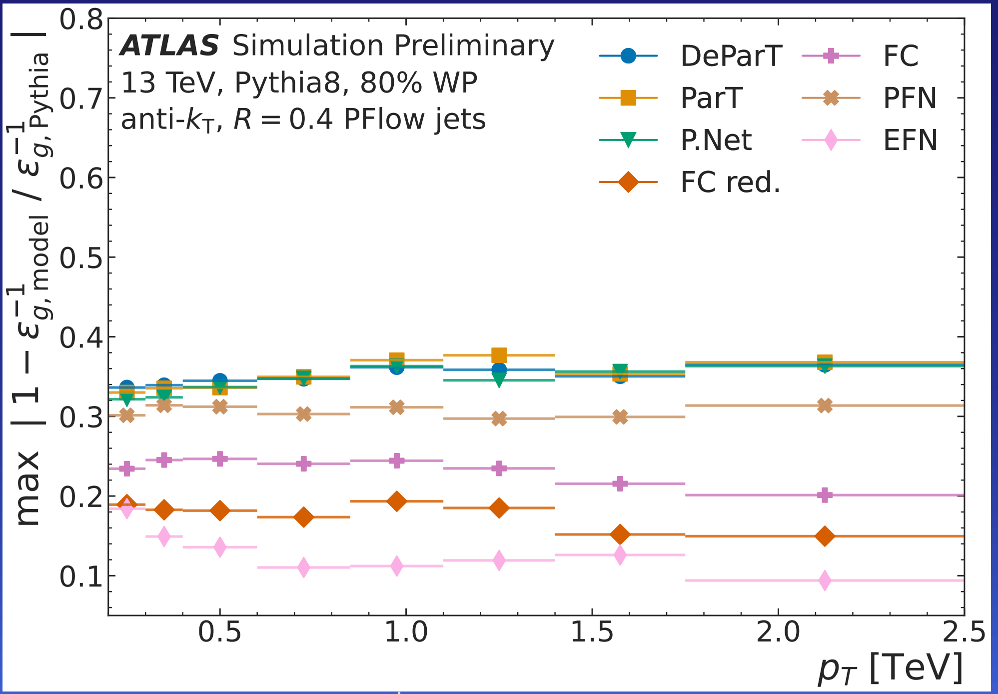
Identification methods

- Distinguishing jets initiated by different particles (light quark, gluon, heavy boson, top-quark) is extremely important to identify the final states
- We use machine learning (with different network architectures) for jet tagging on simulated samples
  - Trained either on high-level jet-based quantities
    - Restricted to infrared/collinear safe observables for some
  - Or with additional information from the jet constituents (flow objects)
- Performance is evaluated by comparing background rejection rate for a given signal efficiency for different taggers
- Model dependence is probed by applying the standard-sample trained tagger on different simulated samples with alternative showering and hadronization modelling

- Standard training is done on fully simulated di-jet MC with Pythia8
  - 10 M $R = 0.4$ anti-k_t PFO jets
- Networks with 10 (5) high-level jet quantities FCN (FC reduced)
- Particle Flow (PFN) and IRC-safe Energy Flow (EFN) Networks with constituent information: 8 for PFN (including mass), 4 linear ones for EFN
- Particle Net (P.Net), a graph NN with each constituent (and its 7 features) forming a node, connected via edges to $k = 16$ nearest neighbours
- Particle Transformer (ParT) and Dynamically Enhanced Particle Transformer (DeParT) with constituent features (like P.Net) and interaction variables on pairs of constituents
- PartT,DePartT and P.Net best in gluon rejection, but largest in model dependence; least model dependence in IRC-safe EFN

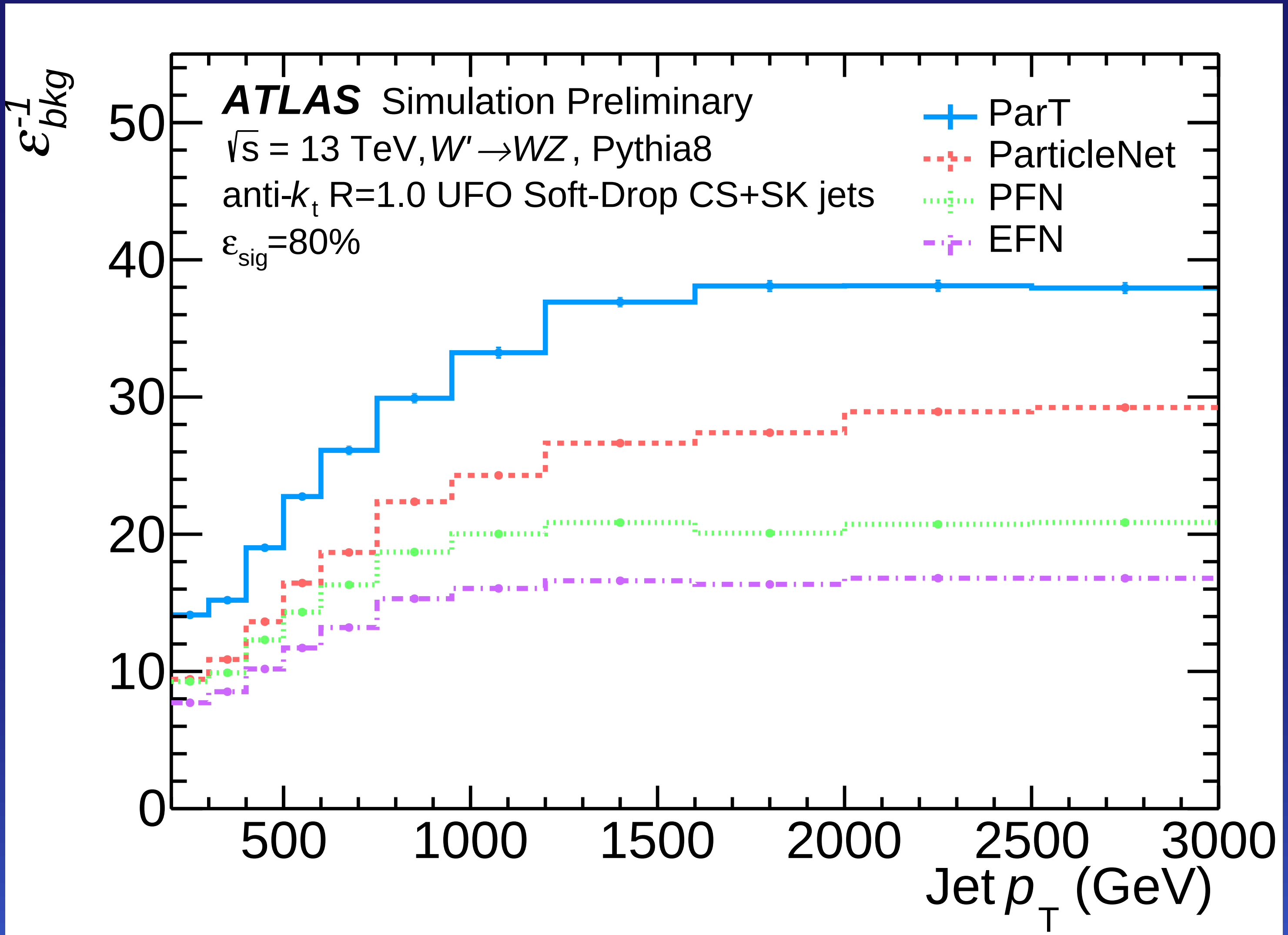


ϵ_g^{-1} vs. $p_\perp$ for $\epsilon_q = 0.8$

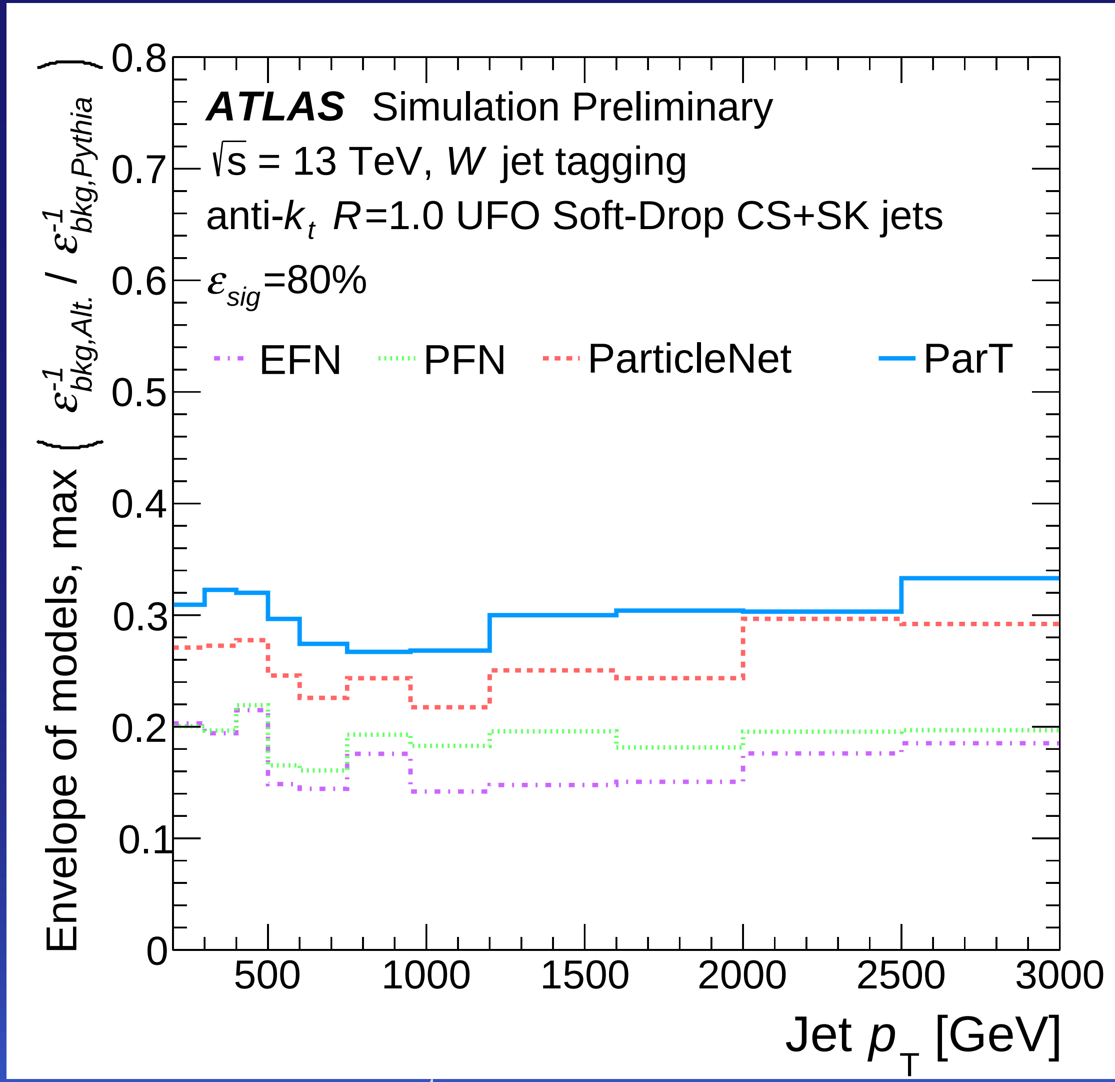


Rel. deviation of ϵ_g^{-1} from std. sample for $\epsilon_q = 0.8$

- PFN, EFN, P.Net and ParT are trained on Pythia8 generated large $R = 1.0$ anti-k_t UFO jets from $W' \to WZ$ events and Pythia8 generated multi-jet background
- ParT with highest bkgd rejection (and largest model dependence), EFN smallest in both

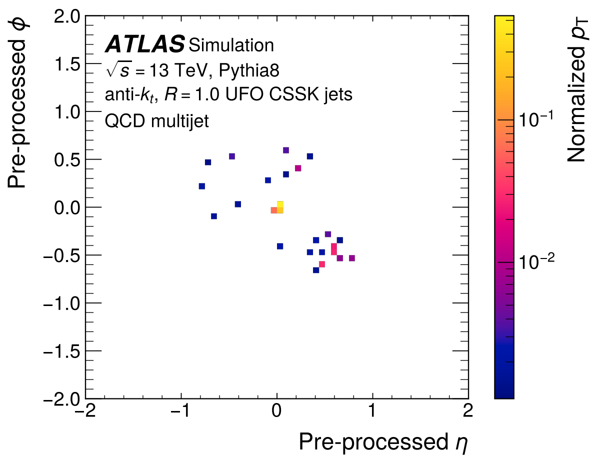


$\epsilon^{-1}_{\mathrm{bkgd}}$ vs. $p_\perp$ for $\epsilon_{\mathrm{sig}} = 0.8$

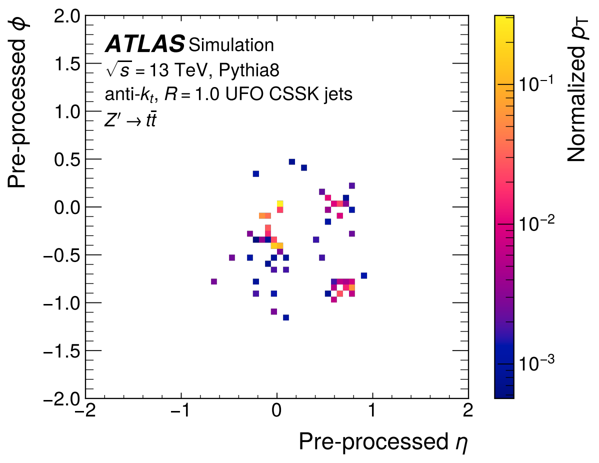


Rel. deviation of $\epsilon^{-1}_{\mathrm{bkgd}}$ from std. sample for $\epsilon_{\mathrm{sig}} = 0.8$

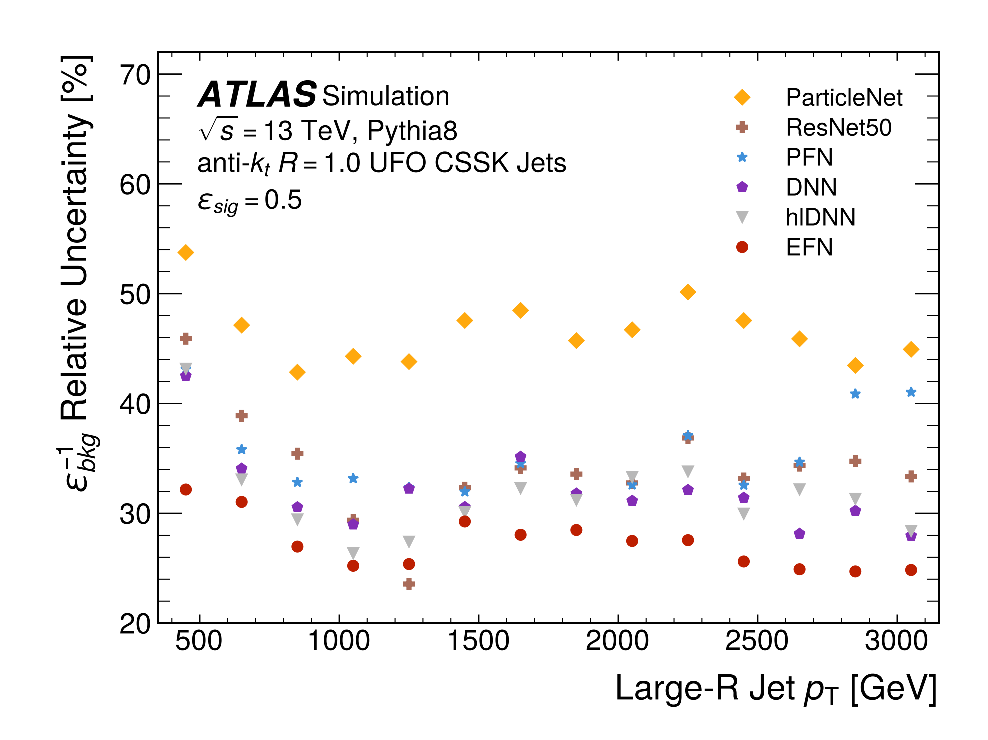
- Baseline hlDNN (high-level-jet info) compared to constituent based DNN, PFN, EFN, P.Net and ResNet50 are trained on Pythia8 generated large $R = 1.0$ anti-k_t UFO jets from $Z' \to t\bar{t}$ events and Pythia8 generated multi-jet background
- ResNet50 is an image classification CNN
  - Turn every jet into a $2D$ image of energies of 64×64 rotated $\Delta\eta \times \Delta\phi$ pixels
- ”Bottom-up” experimental uncertainites evaluated in addition to model-dependence
- P.Net with highest bkgd rejection (and largest model dependence), EFN smallest uncertainty; ResNet50 worst rejection, 2nd largest uncertainty



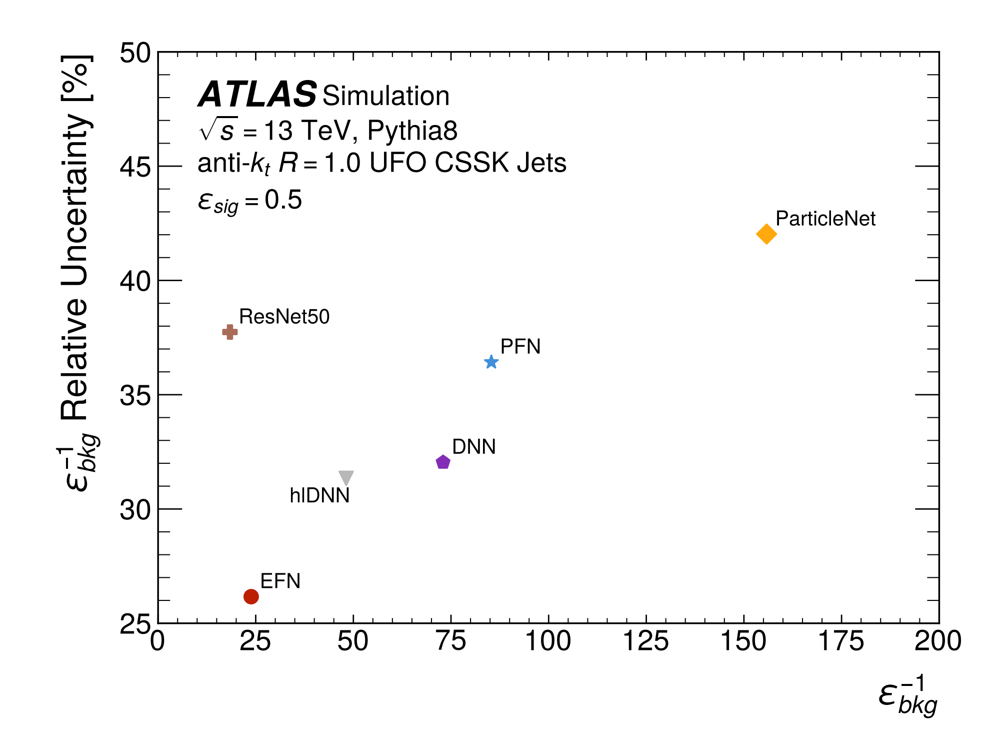
Background jet image



Signal jet image



Rel. uncertainty of $\epsilon_{\mathrm{bkgd}}^{-1}$ for $\epsilon_{\mathrm{sig}} = 0.5$



Rel. uncertainty of $\epsilon_{\mathrm{bkgd}}^{-1}$ vs. $\epsilon_{\mathrm{bkgd}}^{-1}$ for $\epsilon_{\mathrm{sig}} = 0.5$

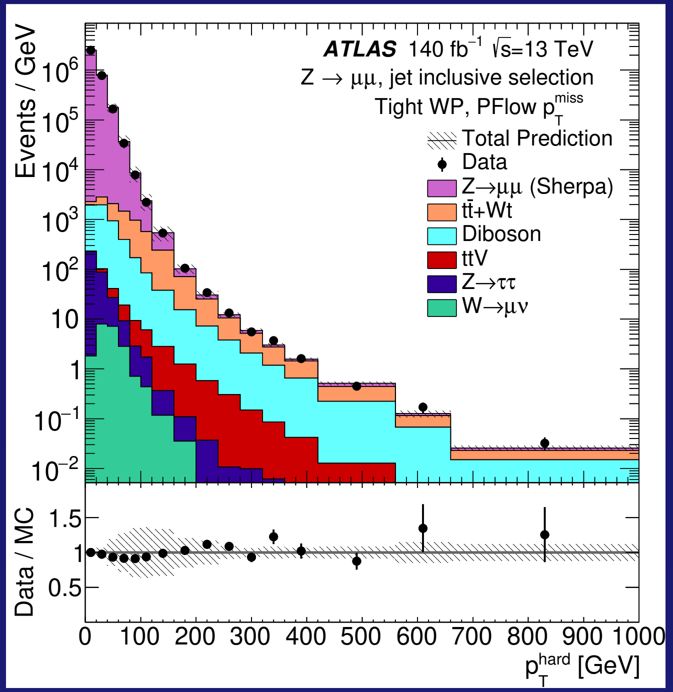
- 2D missing transverse momentum vector $\boldsymbol{p}_\perp^{\text{miss}}$ is derived from $\boldsymbol{p}_\perp^{\text{obj}} = (p_x^{\text{obj}}, p_y^{\text{obj}})$ from all "hard" objects (obj) and a remaining "soft" term from "unused" tracks $\boldsymbol{p}_\perp^{\text{track}}$:

$$\boldsymbol{p}_\perp^{\text{miss}} = -\boldsymbol{p}_\perp^{\text{hard}} - \boldsymbol{p}_\perp^{\text{soft}}, \text{ with } \quad \boldsymbol{p}_\perp^{\text{hard}} = \sum_{\text{obj}=e,\gamma,\tau,\mu,\text{jet}} \boldsymbol{p}_\perp^{\text{obj}} \quad \text{and} \quad \boldsymbol{p}_\perp^{\text{soft}} = \sum_{\text{unused tracks}} \boldsymbol{p}_\perp^{\text{track}}$$

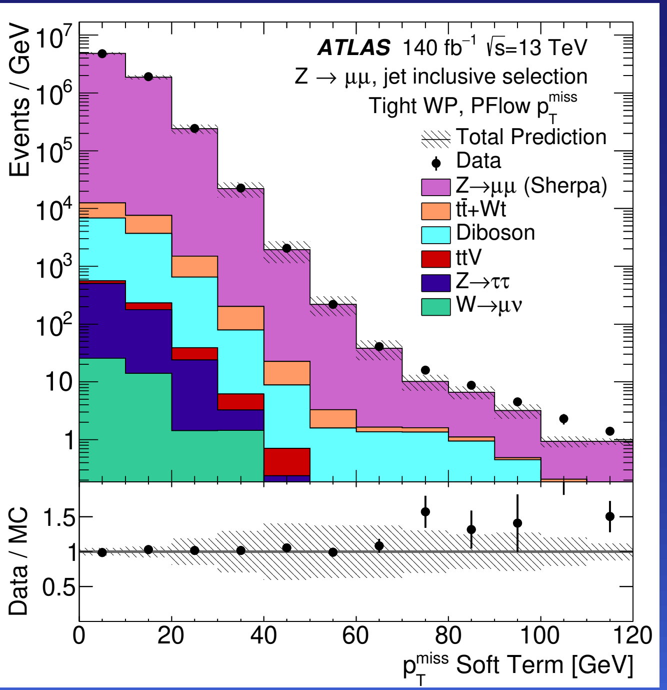
- Scalar transverse momentum sum to evaluate the scale:

$$\sum p_\perp = \sum_{\text{obj}=e,\gamma,\tau,\mu,\text{jet}} p_\perp^{\text{obj}} + \sum_{\text{unused tracks}} p_\perp^{\text{track}}$$

- Run-2 performance updated with full Run-2 dataset for use of PFlow objects for jets
- Evaluation in $Z^0 \to \mu\mu$ and $Z^0 \to ee$ events (no real $p_\perp^{\text{miss}}$ expected)
  - Dominant systematic in $p_\perp^{\text{hard}}$ from JES (bump at $\sim 100\,\text{GeV}$)
  - Small excess in $p_\perp^{\text{soft}}$ tail in data from fake electrons



$p_\perp^{\text{hard}}$ in $Z^0 \to \mu\mu$



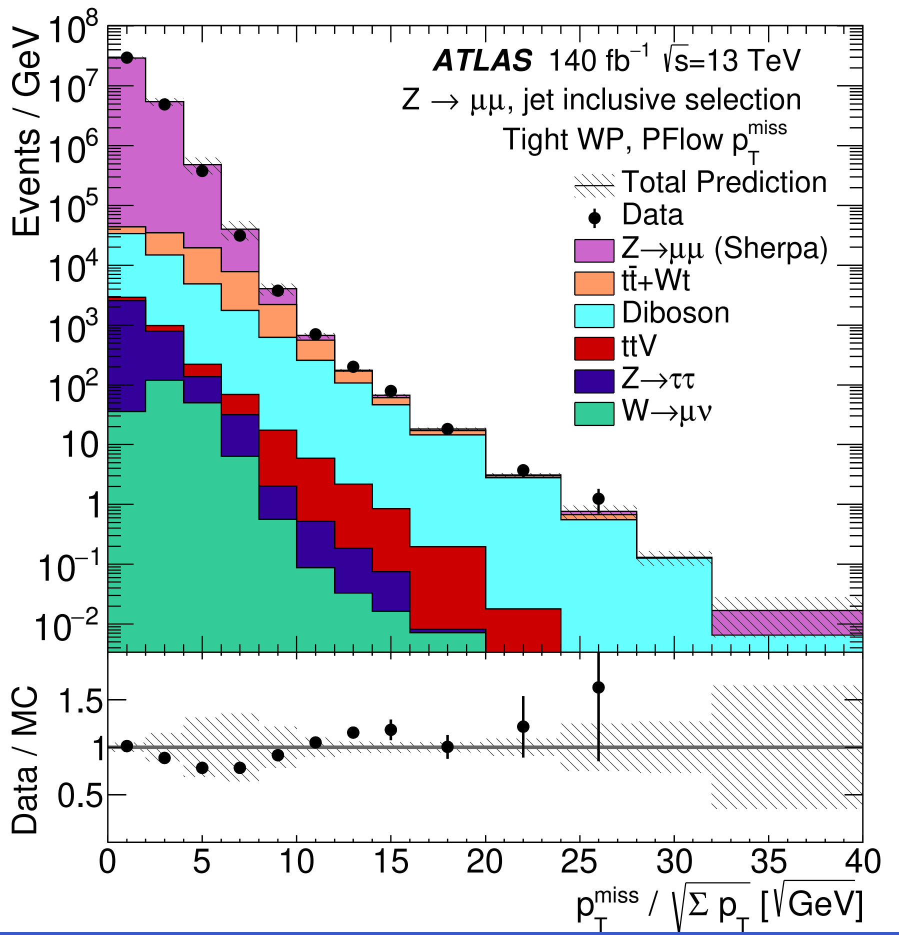
$p_\perp^{\text{soft}}$ in $Z^0 \to \mu\mu$

Missing transverse momentum

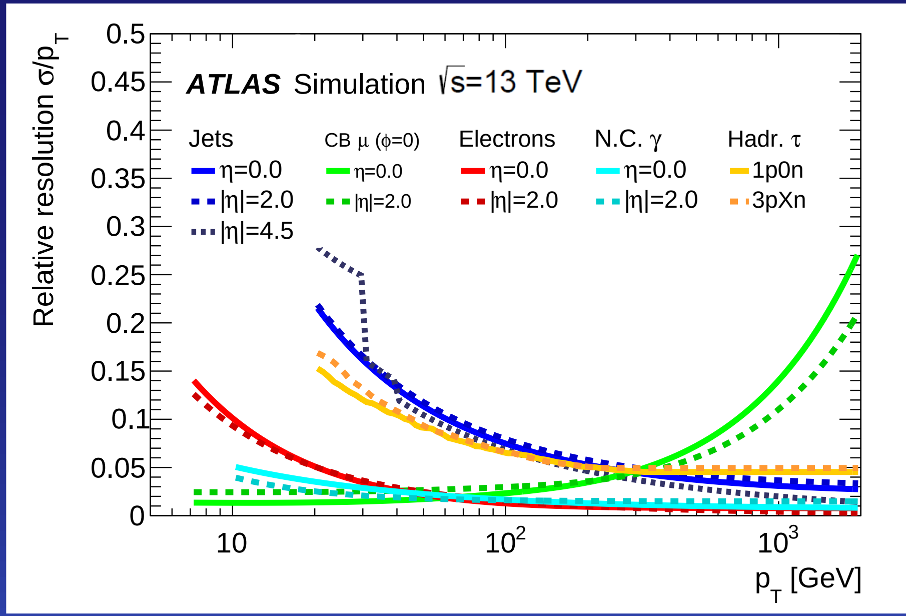
- Event-based significance $S_{H_\perp} = p_\perp^{\text{miss}} / \sqrt{H_\perp}$ is based on $H_\perp = \sum_{\text{jet}} p_\perp^{\text{jet}}$, which is approximate only (assumes calorimeter-like resolution)

- Missing transverse momentum significance evaluated on object-based uncertainties $\boldsymbol{V}$:

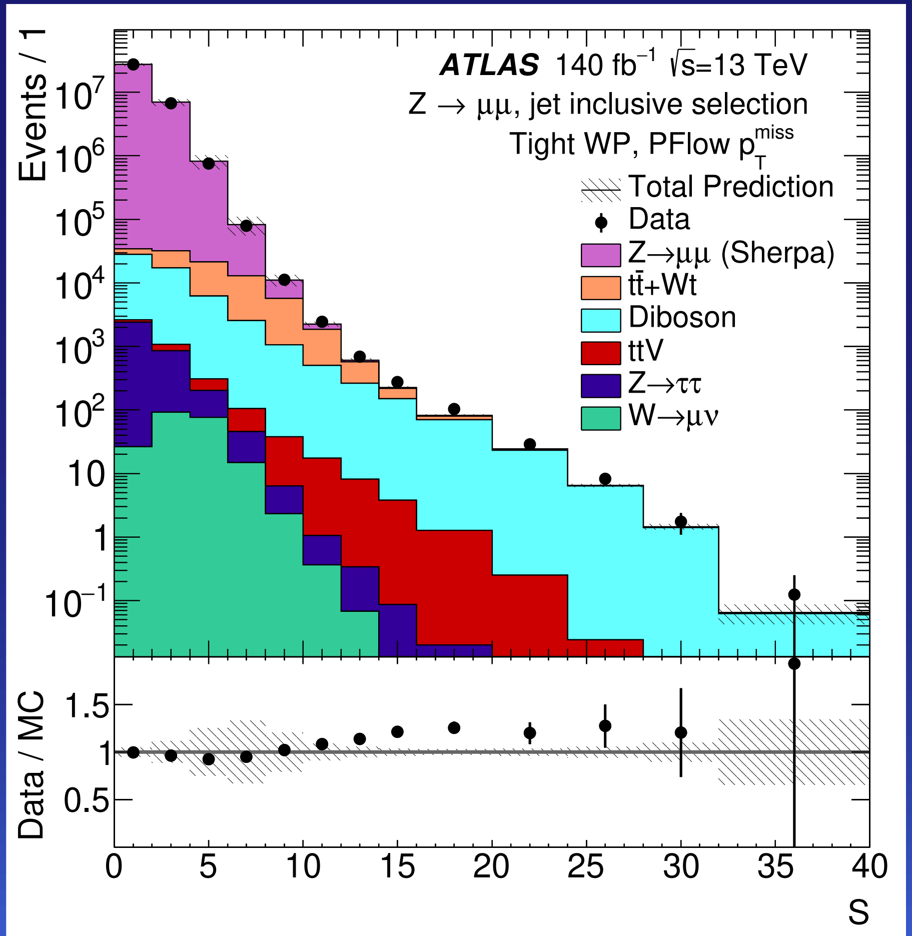
$$S = \sqrt{\boldsymbol{p}_\perp^{\text{miss}\,T} \left(\sum_{\text{obj}} \boldsymbol{V}^{\text{obj}} \right)^{-1} \boldsymbol{p}_\perp^{\text{miss}}}$$



$p_\perp^{\text{miss}} / \sqrt{\sum p_\perp}$ in $Z^0 \to \mu\mu$



Relative resolutions of the objects entering $p_\perp^{\text{miss}}$



Object-based significance S in $Z^0 \to \mu\mu$

Conclusions

- Reconstruction and calibration of hadronic objects in ATLAS is a very active field
  - Pile-Up remains the biggest challenge
  - Time as a new discriminant in calorimetry helps reducing it
  - New ML-based techniques start to replace legacy calibration methods for energy and mass
- Run-2 performance results:
- Jet calibration
  - $O(1\%)$ precision reached for jet energy scale, $O(15 - 30\%)$ improvements in resolution for energy and mass
  - Additional b-jet energy scale uncertainty measured to $O(1\%)$
- Jet tagging
  - q/g, heavy-boson and t-quark taggers based on ML with constituent information outperform taggers with high-level jet info
  - But model dependence is larger for constituent based taggers
- Missing transverse momentum
  - Benefits from reconstruction and calibration advancements – especially from jets
  - Object-based significance sharpens the MET discrimination power
- Run-3 analyses benefit from these improvements