Precision calibration of calorimeter signals in the ATLAS experiment using an uncertainty-aware neural network

arXiv:2412.04370 [hep-ex]
Submitted to SciPost

Jad M. Sardain on behalf of the ATLAS Collaboration
with Theo Heimel, Tilman Plehn, Philip Velie, Lorenz Vogel

BOOST2025, July 28 – August 1, 2025

Overview

- Topo-clusters are formed by grouping calorimeter cells with significant energy deposits that are topologically connected

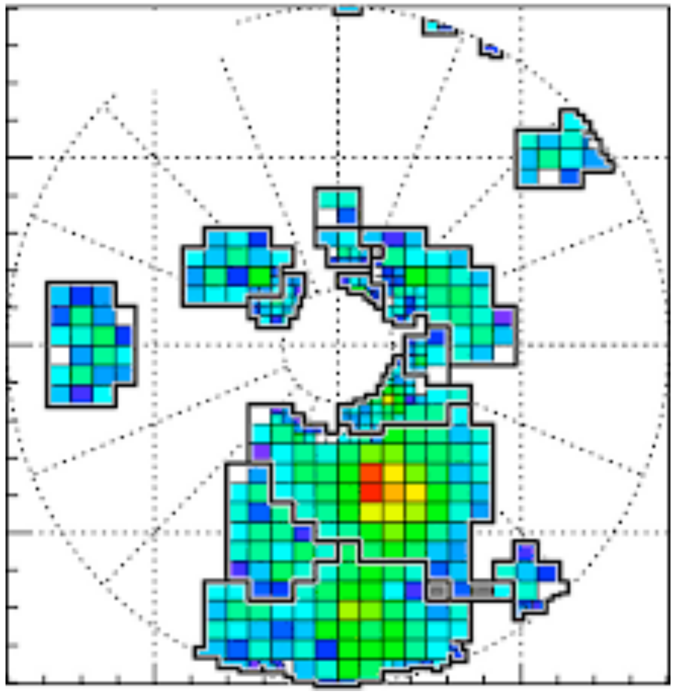
- They represent the **primary signals** used by the calorimeter

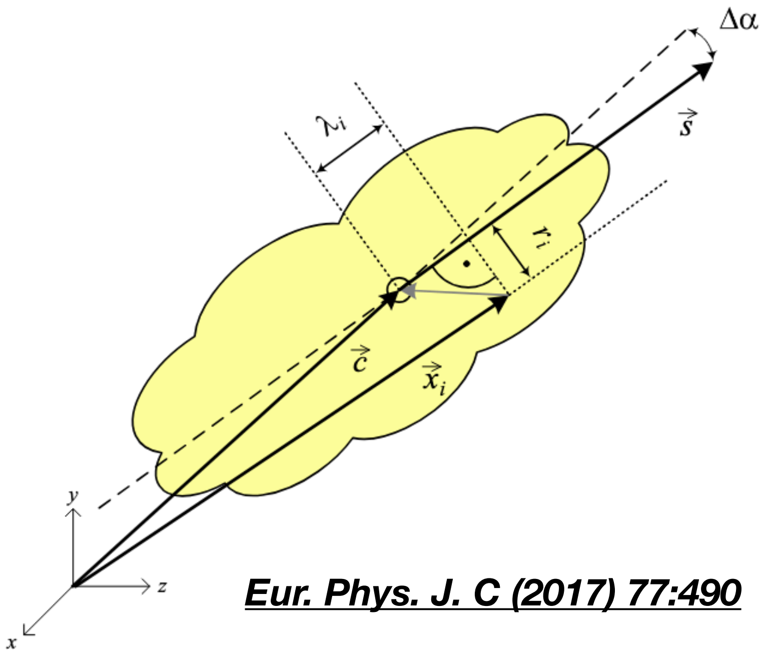
- They are calibrated to **correctly measure** the energy deposited by EM showers

- However, they are **not fully corrected** for all the energy lost in hadronic showers

→ Multi-dimensional correlated calibration

Eur. Phys. J. C (2017) 77:490





Eur. Phys. J. C (2017) 77:490

$\vec{c}$ *centre of gravity of cluster, measured from the nominal vertex* $(x = 0, y = 0, z = 0)$ *in ATLAS*

$\vec{x}_i$ *geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*

$\vec{s}$ *particle direction of flight (shower axis)*

$\Delta\alpha$ *angular distance* $\Delta\alpha = \angle(\vec{c}, \vec{s})$ *between cluster centre of gravity and shower axis* $\vec{s}$

λ_i *distance of cell at* $\vec{x}_i$ *from the cluster centre of gravity measured along shower axis* $\vec{s}$ $(\lambda_i < 0$ *is possible)*

r_i *radial (shortest) distance of cell at* $\vec{x}_i$ *from shower axis* $\vec{s}$ $(r_i \geq 0)$

Overview

- Standard approach (LCW) [Eur. Phys. J. C (2017) 77:490]
  - 4-step sequence where each step employs cluster features as input to a (binned) look-up table retrieving calibration and correction factors
  - *Limitations:* **1)** average factors in multi-dimensional bins **do not consider correlations**
    2) bin boundaries introduce **steep transitions**
    3) can only be re-applied in full data reprocessing (1-2 times per year)

- Machine learning based approach
  - **Regression fit of signal to energy deposited** at topo-cluster location using similar features as LCW
  - **Preserves** (and exploits) **correlations** and **avoids step-like** transition (smooth multi-dimensional functions)
  - Can **accommodate** adaptations, optimizations, and fine-tuning **during run year**
  - Use of pre-processed information

The provided calibration is uncertainty-aware
It does not only calibrate the clusters to the target energy scale, but also provides the per-cluster uncertainty

Machine Learning: Training Goals

Target:

Learn the response $\boxed{\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}} = E^{\mathrm{EM}}_{\mathrm{clus}}/E^{\mathrm{dep}}_{\mathrm{clus}}}$ as a function of a set of observables $\mathcal{X}_{\mathrm{clus}}$ sensitive to the energy deposit

$E^{\mathrm{EM}}_{\mathrm{clus}}$ is the topo-cluster signal at EM scale (basic signal) (reconstructed detector signal)

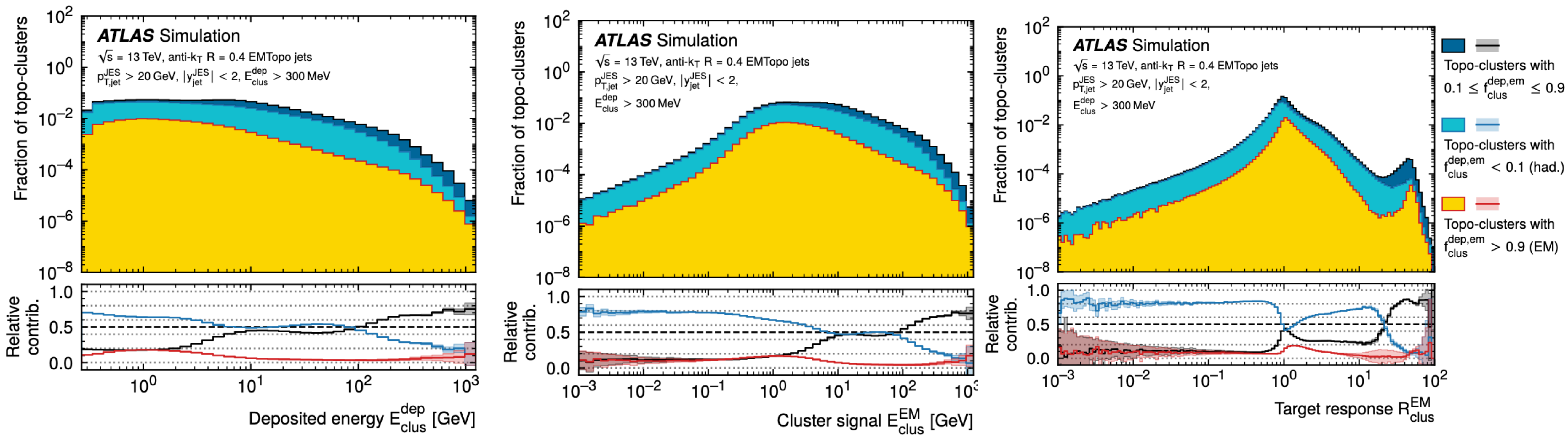
$E^{\mathrm{dep}}_{\mathrm{clus}}$ is the energy deposited in cells forming the cluster (True deposited energy, in MC only)

Prediction:

Successfully trained network predicts response $\boxed{\mathcal{R}^{\mathrm{NN}}_{\mathrm{clus}}(\mathcal{X}_{\mathrm{clus}})}$ per cluster for any phase space point/region

Final calibration is then applied cluster-by-cluster by: $\boxed{E^{\mathrm{NN}}_{\mathrm{clus}} = \dfrac{E^{\mathrm{EM}}_{\mathrm{clus}}}{\mathcal{R}^{\mathrm{NN}}_{\mathrm{clus}}(\mathcal{X}_{\mathrm{clus}})}}$

Target has better dynamic range than targeting $E^{\mathrm{dep}}_{\mathrm{clus}}$ directly



Samples and input features

- The ML based topo-cluster calibration is derived using fully simulated dijet MC events
  - **Realistic** Run-2 collision environment **with pileup,** different processes included (hard scattering, UE, etc.)

- All topo-clusters used are required to:
  - be part of a calorimeter jet, anti-k_t $R = 0.4$ jets with $p_{T,\text{jet}} > 20$ GeV, $|y_\text{jet}| < 2.0$
  - have at least partial real energy signal $E^\text{dep}_\text{clus} > 300$ MeV
  - Pure pile-up clusters are rejected as they will have $E^\text{dep}_\text{clus} = 0$ yielding to undefined response

- Resulting sample has $14 \cdot 10^6$ clusters

- Input features: $\boxed{\mathcal{X}_\text{clus}}$

Category	Symbol	Comment
Kinematics	E^EM_clus	Signal at the electromagnetic energy scale
	y^EM_clus	Rapidity at the electromagnetic energy scale
Signal strength, time structure	$\zeta^\text{EM}_\text{clus}$	Signal significance
	$\text{Var}_\text{clus}(t_\text{cell})$	Variance of t_cell distribution
Shower depth, shower shape, compactness	λ_clus	Distance of the centre-of-gravity from the calorimeter front face (along principal cluster axis)
	$\lvert\vec{c}_\text{clus}\rvert$	Distance of the centre-of-gravity from the nominal vertex
	f_emc	Fraction of energy in the electromagnetic calorimeter
	$\langle\rho_\text{cell}\rangle$	Cluster signal density measure
	$\langle\mathfrak{m}^2_\text{long}\rangle$	Normalised energy dispersion along the principal cluster axis
	$\langle\mathfrak{m}^2_\text{lat}\rangle$	Normalised energy dispersion perpendicular to the principal cluster axis
	$p_\text{T}D$	Signal compactness measure (inspired by Ref. [30])
Topology	f_iso	Cluster isolation measure
Pile-up	t_clus	Signal timing
	N_PV	Number of reconstructed primary vertices
	μ	Number of pile-up interactions per bunch crossing

Wide coverage in topo-cluster phase space

Feature selections is guided by:
- Expressiveness for the response
- Pile-up sensitivity
- Features in data are well modeled in MC

Mismodeling can lead to uncontrollable effects when highly dimensional calibration models are applied to data

Models used

- "Simple" DNN have been studied ***ATL-PHYS-PUB-2023-019*** and can be used as benchmark

- Unlike usual MLPs, Bayesian Neural Networks **replace** the **weights** with a weight distribution $q(\theta)$

- We do not **just** learn the response

- The predicted response is then constructed from the N sampled predictions

- Uncertainties can also be derived from the networks. They are divided into two categories: *arXiv:1904.10004 [hep-ph]*
    - *epistemic* or *model uncertainties*: related to how machines can learn and what cannot be learned (~stat.)
    - *aleatoric* or *data uncertainties*: related to the residual level of stochasticity in the data that cannot be mitigated any further (~syst.)

- With sufficiently large training dataset, $\sigma_{stat} \rightarrow 0$, and $\sigma_{syst} \rightarrow$ const. > 0

- This value depends on the presence of:
    1) non-deterministic or stochastic target (noisy data) which is the case of $\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}$ *arXiv:2003.11099 [hep-ph]*
    2) limited expressiveness of the network (structure uncertainty) *arXiv:2206.14831 [hep-ph]*
    3) non-optimal choices of hyper-parameters (uncertainty in model parameters)

- The uncertainty predictions derived from the *weight space sampling* (BNN) are **validated** with a *function space sampling* using Repulsive Ensemble (RE)

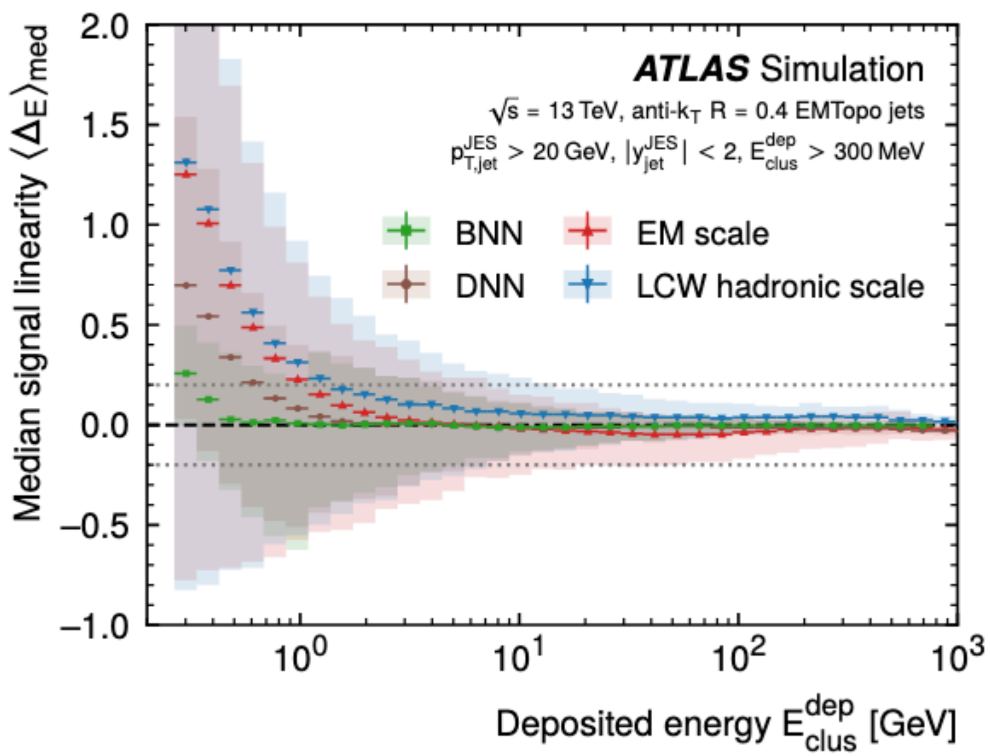
- The sampling is implemented in the loss function rather than in the weight distributions

Results

Two metrics were used to evaluate the performance

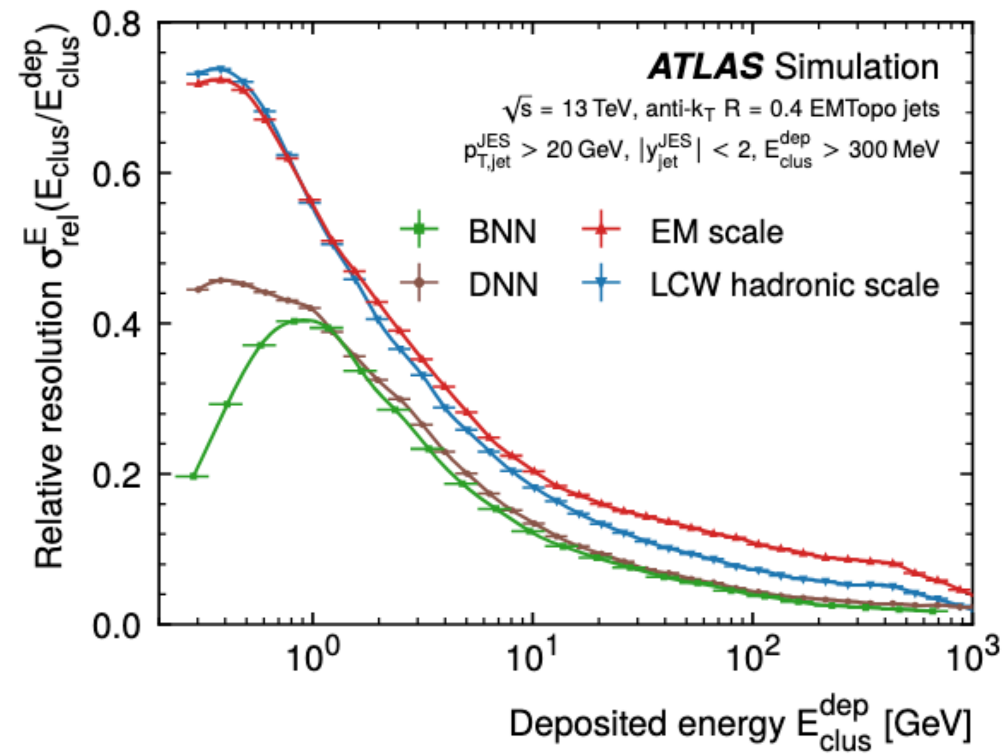
Signal linearity

$$\langle \Delta_E^{\kappa} \rangle_{\text{med}} = \left\langle \left(E_{\text{clus}}^{\kappa} - E_{\text{clus}}^{\text{dep}} \right) / E_{\text{clus}}^{\text{dep}} \right\rangle_{\text{med}}$$



Local topo-cluster energy resolution

$$\text{IQR}_{f=68\%} / \left(2 \times \left\langle E_{\text{clus}}^{\kappa} / E_{\text{clus}}^{\text{dep}} \right\rangle_{\text{med}} \right)$$



- Across the whole energy spectrum, **BNN** gives the **best result**, especially compared to the calibrated LCW

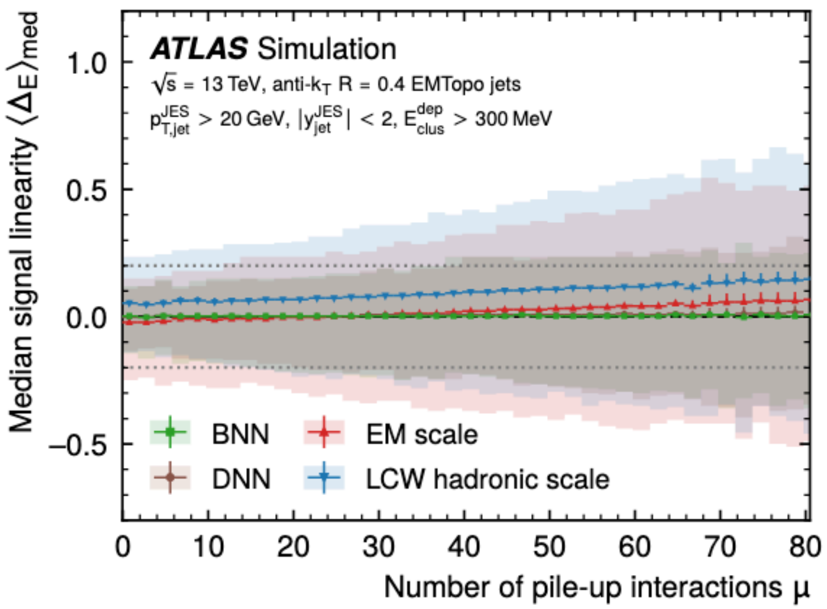
- It performs the best in the low energy regime

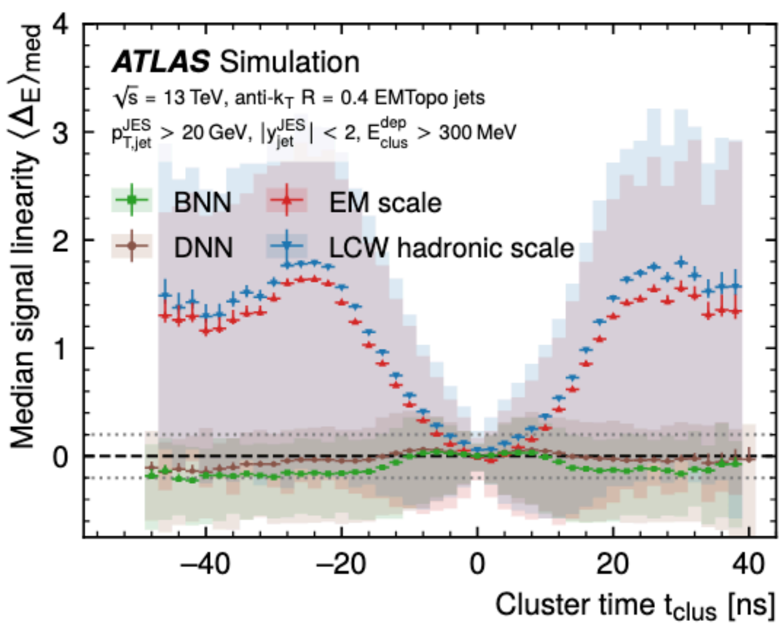
- When comparing the resolution, the BNN tracks changes in the response characteristics well when moving from a **shower-based** signal formation to **ionizations** at low energies

Two metrics were used to evaluate the performance

Signal linearity

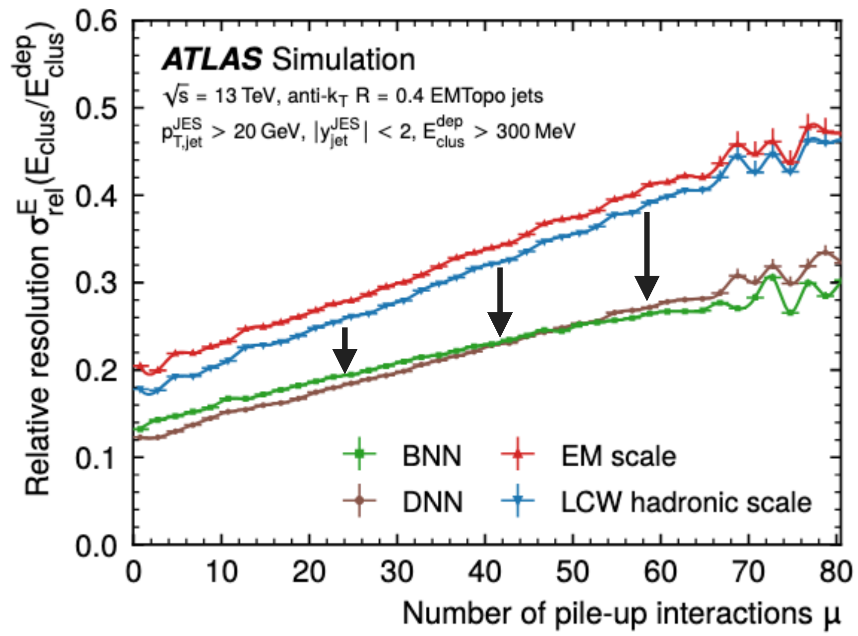
$$\langle \Delta_E^{\kappa} \rangle_{\text{med}} = \left\langle \left(E_{\text{clus}}^{\kappa} - E_{\text{clus}}^{\text{dep}} \right) / E_{\text{clus}}^{\text{dep}} \right\rangle_{\text{med}}$$

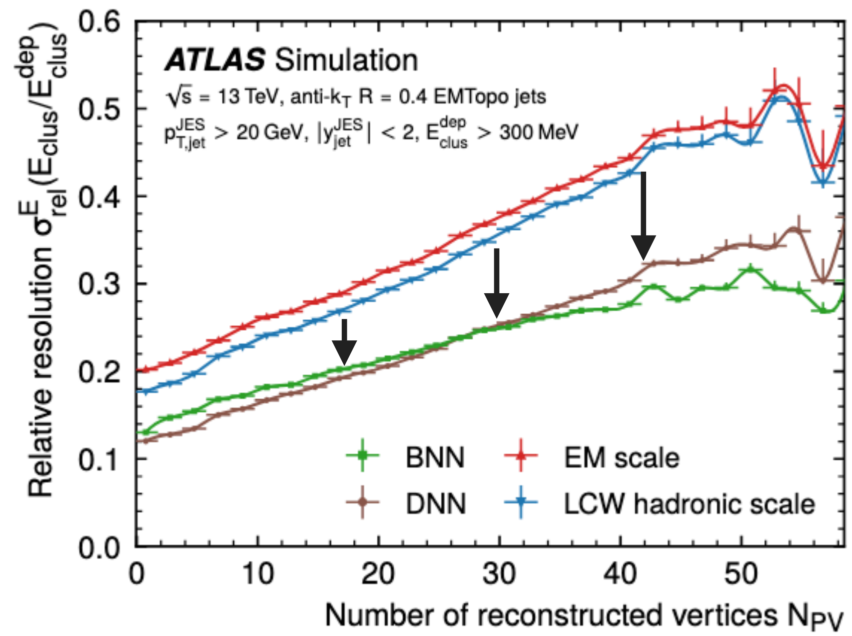




Local topo-cluster energy resolution

$$\text{IQR}_{f=68\%} / \left(2 \times \left\langle E_{\text{clus}}^{\kappa} / E_{\text{clus}}^{\text{dep}} \right\rangle_{\text{med}} \right)$$





- Calibrations is also **environment sensitive** with a significant improvement to the resolution

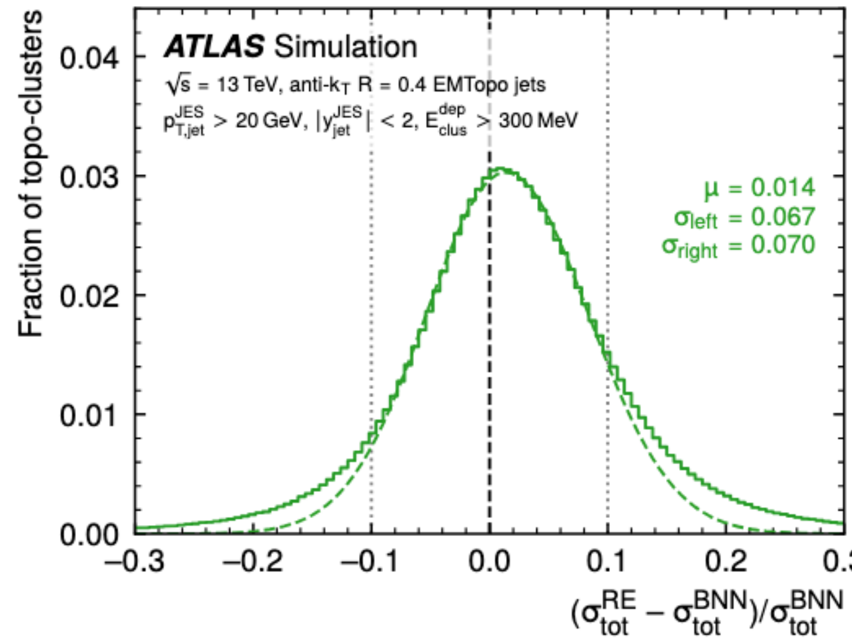
- Strong indication of **pile-up mitigation** of clusters signals by ML calibrations

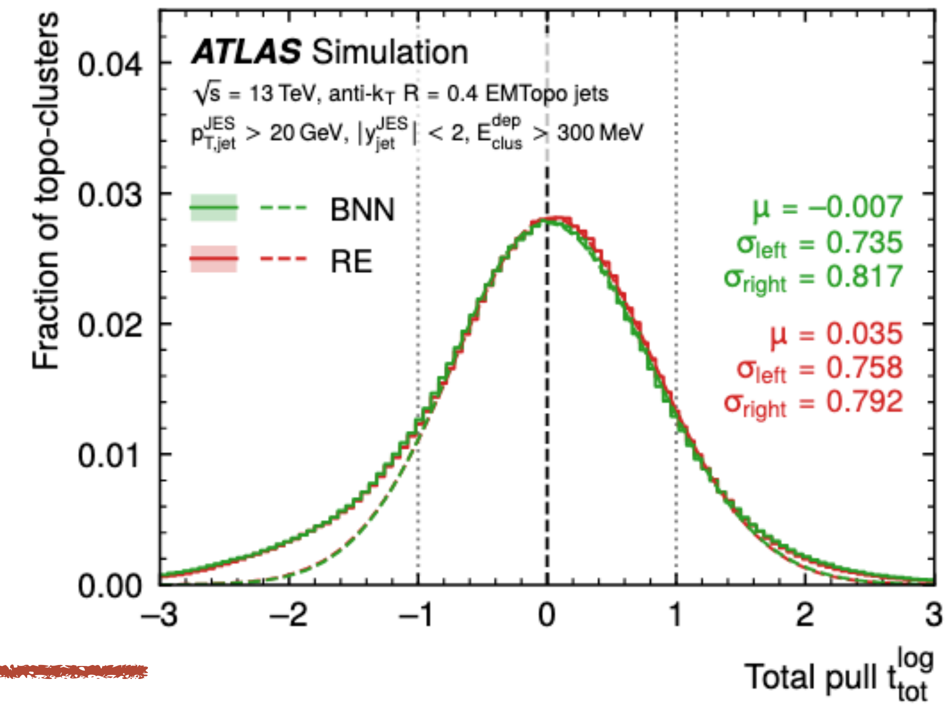
Results: more on uncertainties

- Repulsive Ensemble: force ensemble of identically configured networks to not predict same best-fit parameters

- Resulting uncertainty **very similar** to those of BNN (10% agreement)

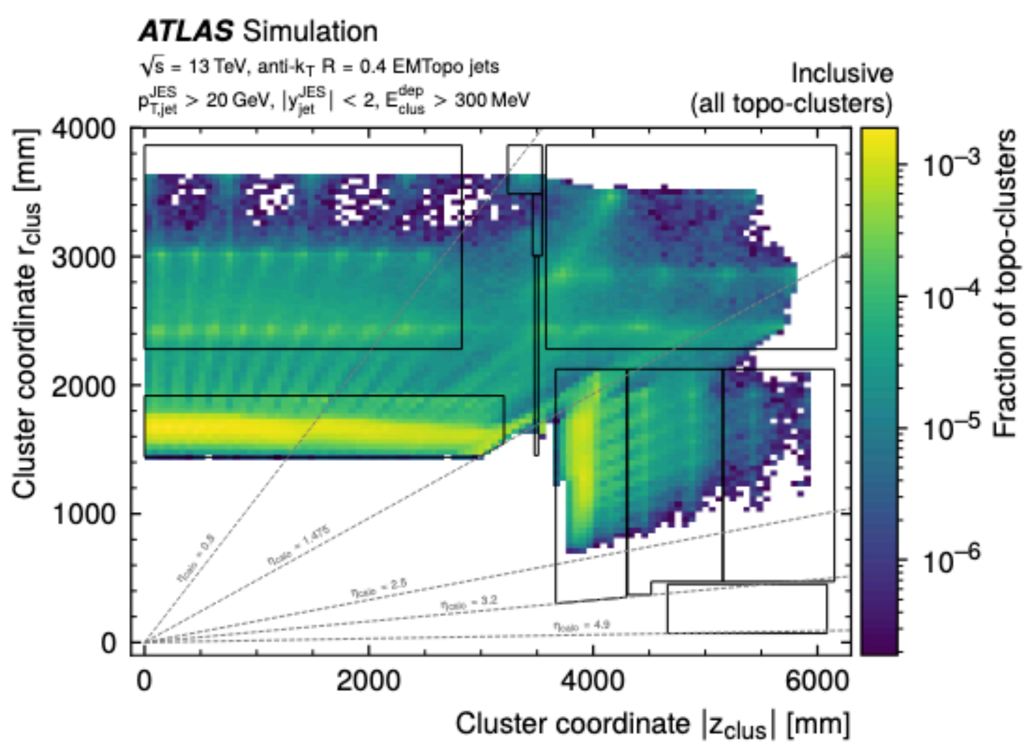
- Both **networks** provide **near identical pulls** cluster-by-cluster (both methods show no bias in their calibrated energy)

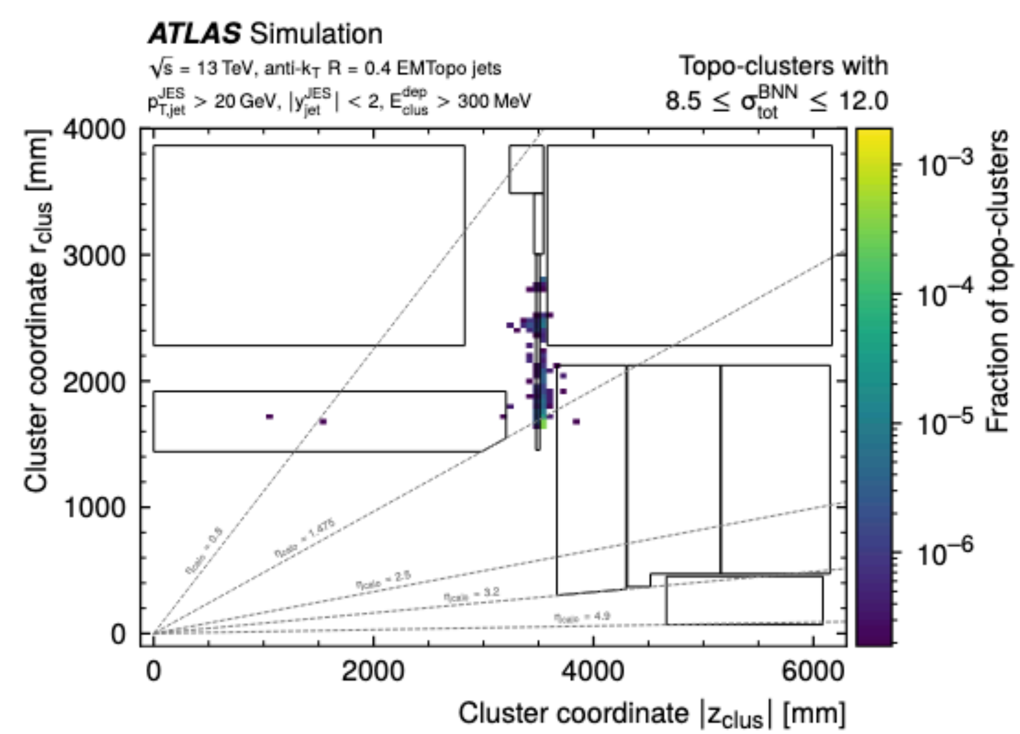




- BNN-uncertainties can also **reveal properties** of the input data

- Plotting location of clusters with **very large calibration uncertainty** show they tend to have significant energy in the **TileGap** region (scintillator)





Conclusions

- **Uncertainty-aware networks** not only provide a **calibration**, but also an **uncertainty** on prediction

- ML based approach is **promising technique** for cluster calibration in a **realistic** collision environment

- Provides an unbinned and continuous calibration

- Better performance compared to standard ATLAS approach, especially in the low energy spectrum

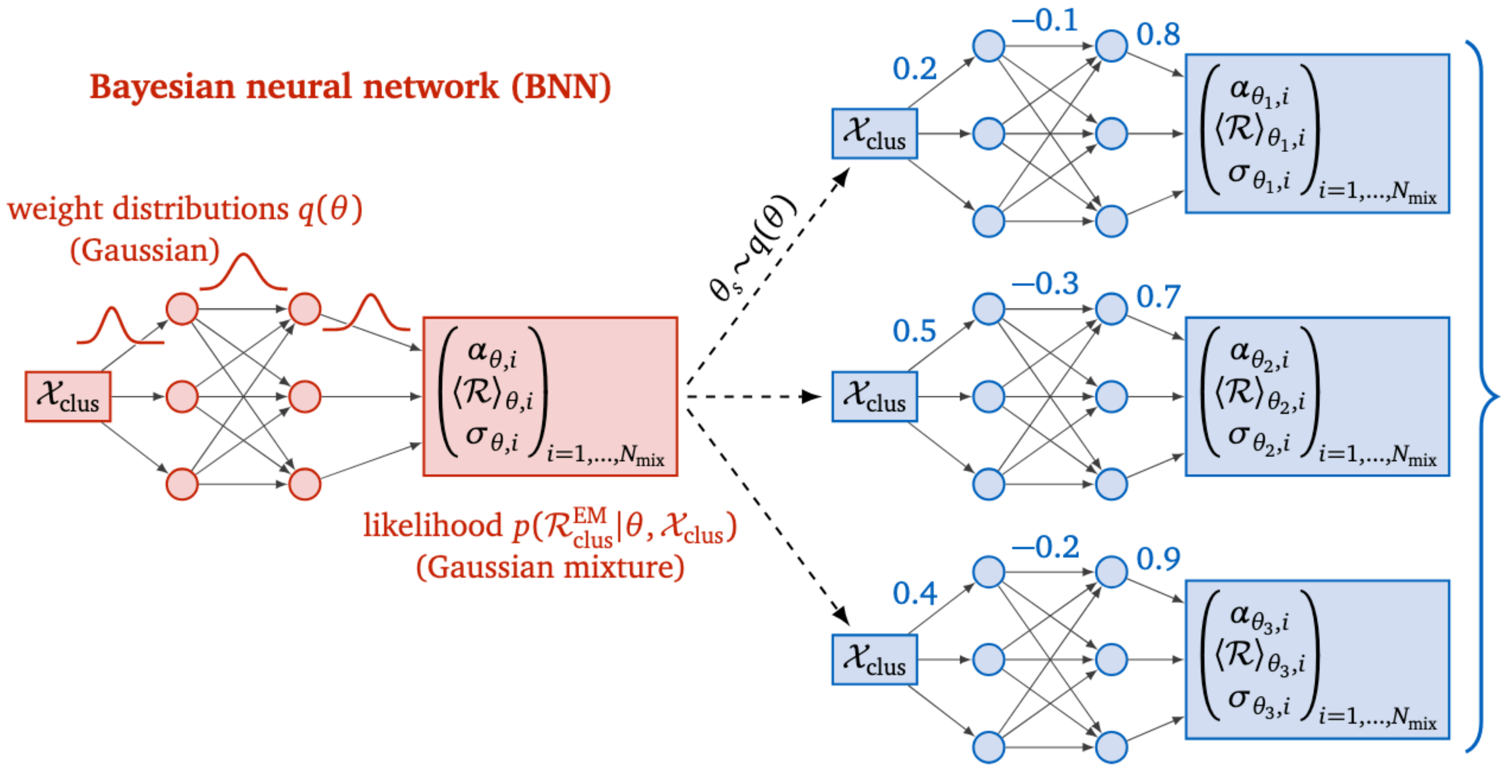
- Bayesian neural networks help also derive per-cluster uncertainties which was never done before

- Networks are also sensitive to, and can pick up, pile-up effects

- **Next big task:** use uncertainty-aware networks with real data. Exciting times ahead!

Back up

Bayesian Neural Network



Central-value prediction (maximum likelihood) and uncertainties for a Gaussian mixture model:

$$\mathcal{R}^{\text{BNN}}_{\text{clus}}(\mathcal{X}_{\text{clus}}) = \arg\max_{\mathcal{R}} \frac{1}{N}\sum_{s=1}^{N} p(\mathcal{R}|\theta_s,\mathcal{X}_{\text{clus}})$$

$$\sigma^2_{\text{syst}}(\mathcal{X}_{\text{clus}}) = \frac{1}{N}\sum_{s=1}^{N}\left[\sum_{i=1}^{N_{\text{mix}}}\alpha_{\theta_s,i}\left(\sigma^2_{\theta_s,i}+\langle\mathcal{R}\rangle^2_{\theta_s,i}\right)-\langle\mathcal{R}\rangle^2_{\theta_s}\right]$$

$$\sigma^2_{\text{stat}}(\mathcal{X}_{\text{clus}}) = \text{Var}\left(\langle\mathcal{R}\rangle_{\theta_s}\right) = \frac{1}{N}\sum_{s=1}^{N}\left[\langle\mathcal{R}\rangle-\langle\mathcal{R}\rangle_{\theta_s}\right]^2$$

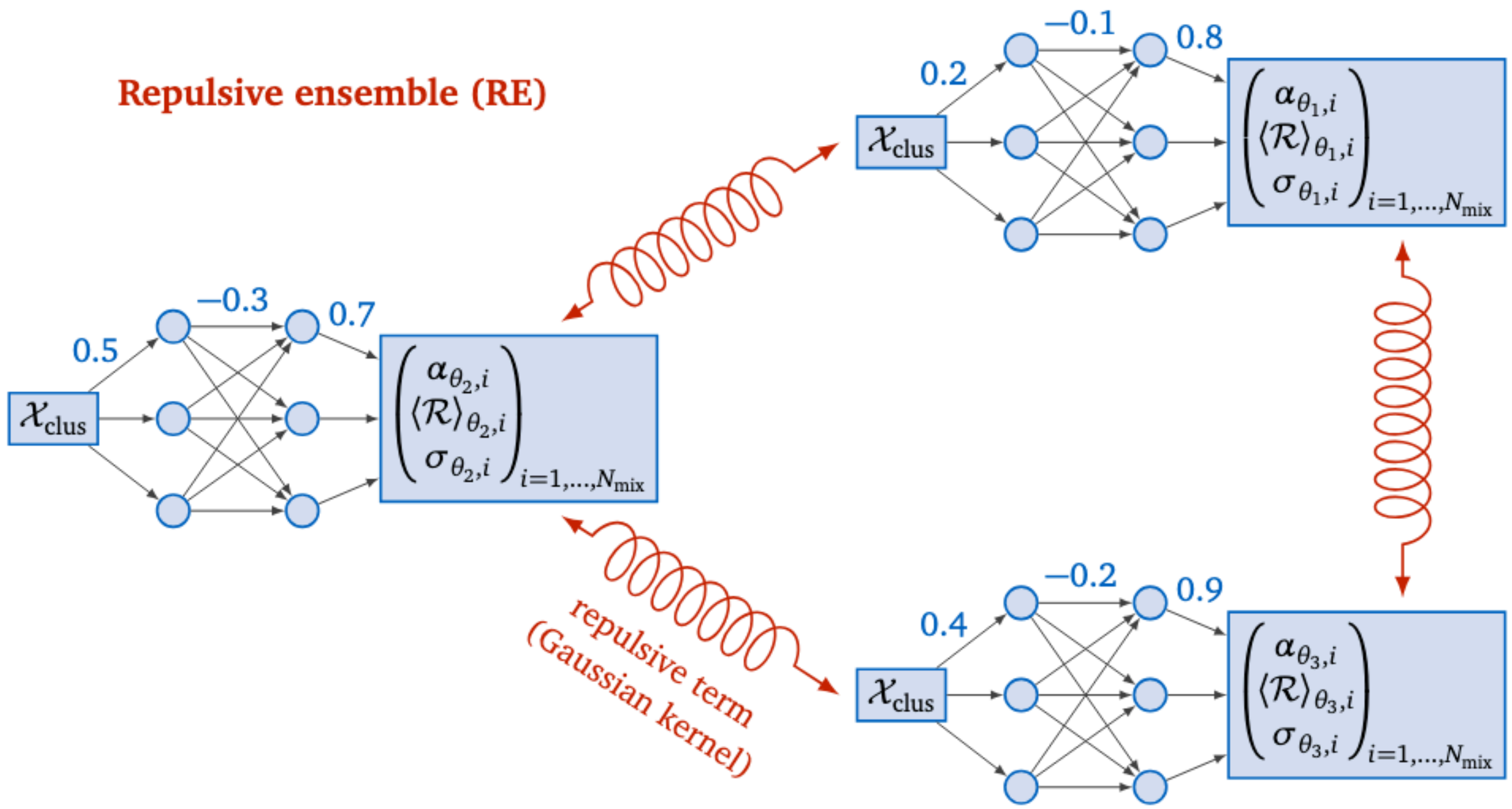
with

$$\langle\mathcal{R}\rangle_{\theta_s} = \sum_{i=1}^{N_{\text{mix}}}\alpha_{\theta_s,i}\langle\mathcal{R}\rangle_{\theta_s,i} \quad \text{and} \quad \langle\mathcal{R}\rangle = \frac{1}{N}\sum_{s=1}^{N}\langle\mathcal{R}\rangle_{\theta_s}$$

Training: Weights linking the nodes of adjacent layers are described by weight distributions $q(\theta)$

Inference: Learned weight distributions $q(\theta)$ are sampled N times to generate a set of network parameters θ_s and thus an ensemble of networks

Repulsive ensemble



Central-value prediction (maximum likelihood) and **uncertainties** for a Gaussian mixture model:

$$\mathcal{R}^{\text{RE}}_{\text{clus}}(\mathcal{X}_{\text{clus}}) = \underset{\mathcal{R}}{\arg\max}\, \frac{1}{N}\sum_{s=1}^{N} p(\mathcal{R}|\theta_s, \mathcal{X}_{\text{clus}})$$

$$\sigma^2_{\text{syst}}(\mathcal{X}_{\text{clus}}) = \frac{1}{N}\sum_{s=1}^{N}\left[\sum_{i=1}^{N_{\text{mix}}} \alpha_{\theta_s,i}\left(\sigma^2_{\theta_s,i} + \langle\mathcal{R}\rangle^2_{\theta_s,i}\right) - \langle\mathcal{R}\rangle^2_{\theta_s}\right]$$

$$\sigma^2_{\text{stat}}(\mathcal{X}_{\text{clus}}) = \text{Var}\left(\langle\mathcal{R}\rangle_{\theta_s}\right) = \frac{1}{N}\sum_{s=1}^{N}\left[\langle\mathcal{R}\rangle - \langle\mathcal{R}\rangle_{\theta_s}\right]^2$$

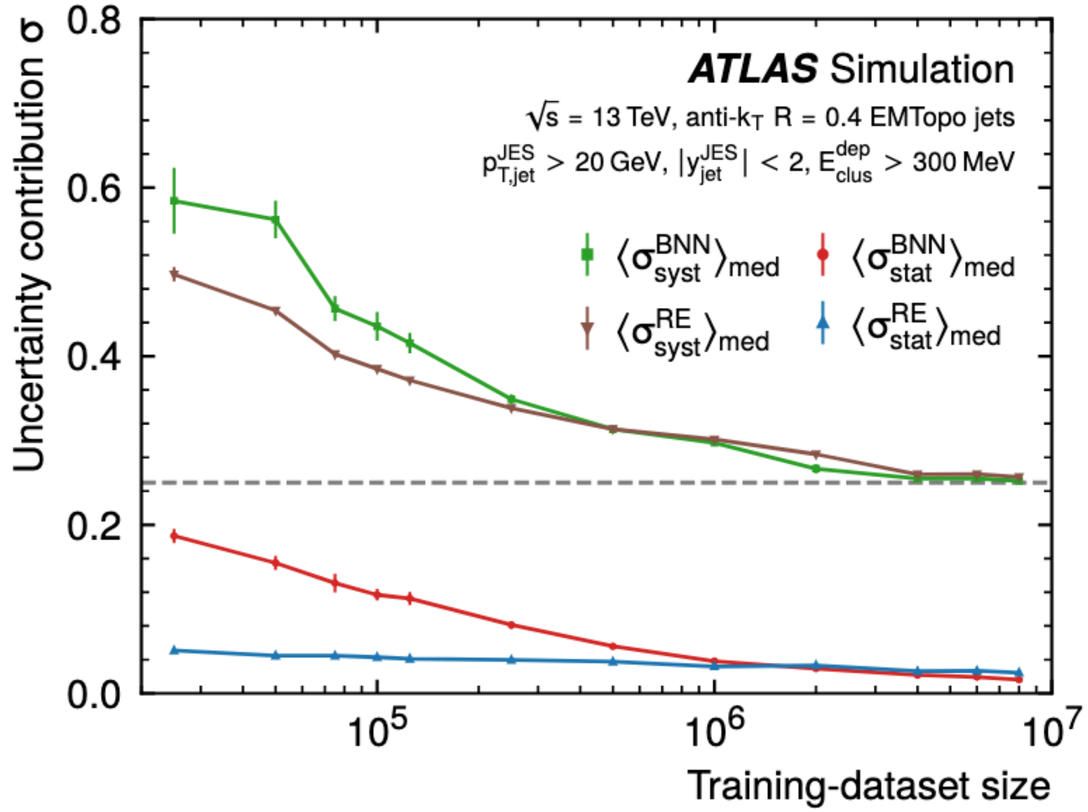
with

$$\langle\mathcal{R}\rangle_{\theta_s} = \sum_{i=1}^{N_{\text{mix}}} \alpha_{\theta_s,i}\langle\mathcal{R}\rangle_{\theta_s,i} \quad \text{and} \quad \langle\mathcal{R}\rangle = \frac{1}{N}\sum_{s=1}^{N}\langle\mathcal{R}\rangle_{\theta_s}$$

Training: Repulsive term connecting the function space of all N simultaneously trained networks forces the ensemble to spread out and cover the loss around the actual minimum

Inference: Same formulas as for the BNN, using the N simultaneously trained ensemble members

Uncertainties



- With sufficiently large training dataset, $\sigma_{stat} \to 0$, and $\sigma_{syst} \to$ const. > 0

- This value depends on the presence of:
  1) non-deterministic or stochastic target (noisy data) which is the case of $\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}$ arXiv:2003.11099 [hep-ph]
  2) limited expressiveness of the network (structure uncertainty) arXiv:2206.14831 [hep-ph]
  3) non-optimal choices of hyper-parameters (uncertainty in model parameters)