Uncertainty-aware BNNs for topo-cluster calibration

What theorists can teach experimentalists about calorimeter calibration

Lorenz Vogel

June 18, 2025

Institute for Theoretical Physics
Heidelberg University, Germany

Intelligent Systems Research Centre (ISRC), Ulster University, UK





Precision calibration of calorimeter signals in the ATLAS experiment using an uncertainty-aware neural network, arXiv:2412.04370 [hep-ex]
ATLAS Collaboration (including T. Heimel, P. Loch, T. Plehn and L. Vogel)

Today's outline



1. CERN and the Large Hadron Collider (LHC)
2. ATLAS detector and calorimeter system
3. Calorimeter-signal calibration (topo-clusters)
4. Bayesian neural networks (BNNs)
5. BNN-calibration performance
6. BNN-learned uncertainties
7. Repulsive ensembles (REs)
8. Summary and outlook

CERN and the Large Hadron Collider

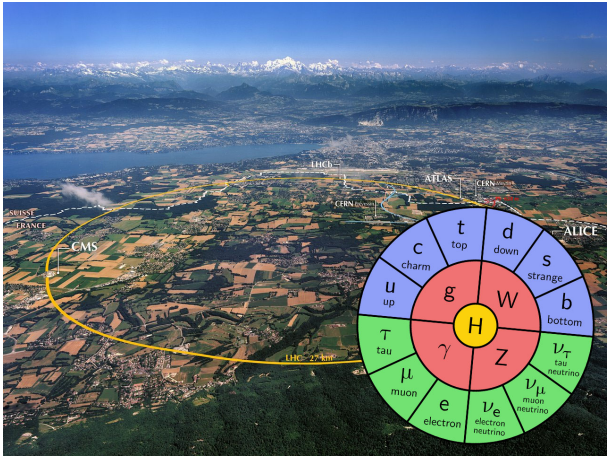
CERN — European Organization for Nuclear Research



Largest **particle-physics laboratory** in the world, founded in 1954

- French-Swiss border near Geneva, 24 member states
- probing basic constituents of matter (**fundamental particles**) using world's largest and most complex scientific instruments
- particle accelerators (LHC) and four main experiments/detectors (ATLAS, CMS, ALICE, and LHCb)

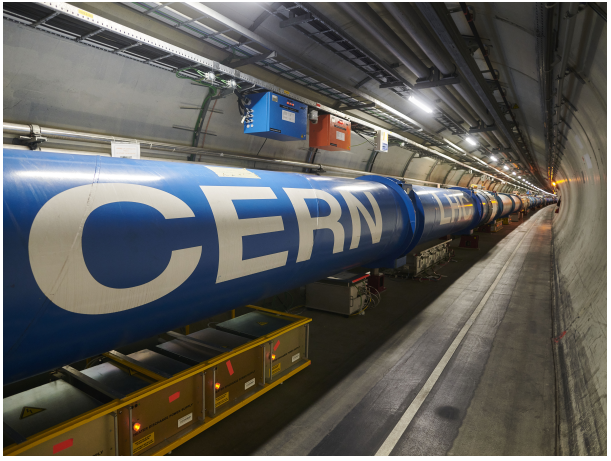
CERN — European Organization for Nuclear Research



Largest **particle-physics laboratory** in the world, founded in 1954

- accelerators boost beams of subatomic particles (protons) to high energies before beams collide with each other — detectors observe and record results of collisions
- study Standard Model (SM) of particle physics and events involving novel physics processes
- SM describes matter at the most fundamental level studied yet

CERN's LHC — Large Hadron Collider



LHC is world's largest and most powerful **particle accelerator**, operational since 2008

- 27 km long accelerator ring, 100 m underground
- two **proton beams** travel at near light-speed in opposite directions through separate ultra-high vacuum beam pipes

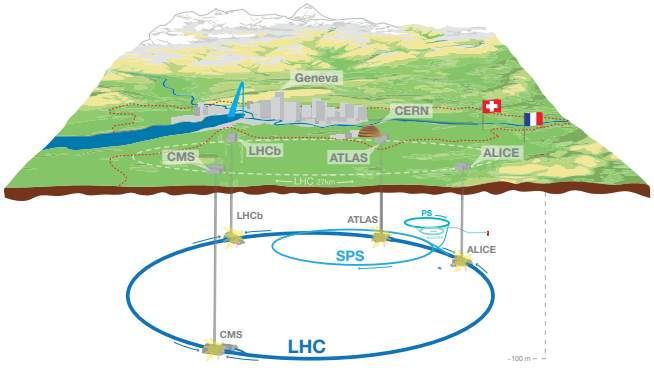
CERN's LHC — Large Hadron Collider



LHC is world's largest and most powerful **particle accelerator**, operational since 2008

- beams guided by strong magnetic fields from superconducting electromagnets (−271.3 °C)
- 1232 dipole magnets (15 m long) to bend beams
- 392 quadrupole magnets (5 m to 7 m) to focus beams

CERN's LHC — Large Hadron Collider

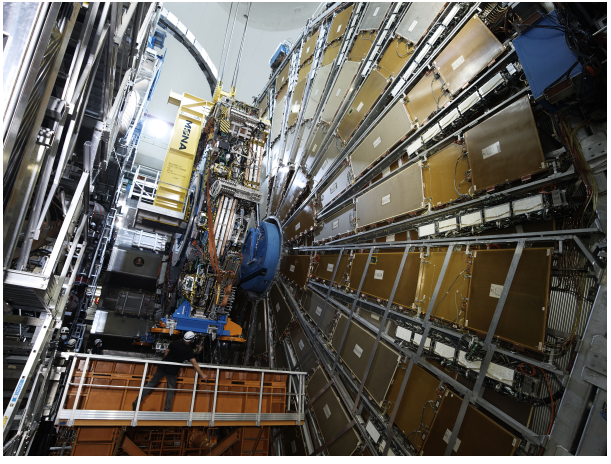


High-energy proton beams **collide at four points** around LHC accelerator ring

- high-energy collisions create new unstable particles, which decay and produce collimated sprays of outgoing particles (jets)
- complex detector systems measure particle properties
- **general-purpose detectors ATLAS and CMS** (Higgs discovery 2012)

ATLAS detector and calorimeter system

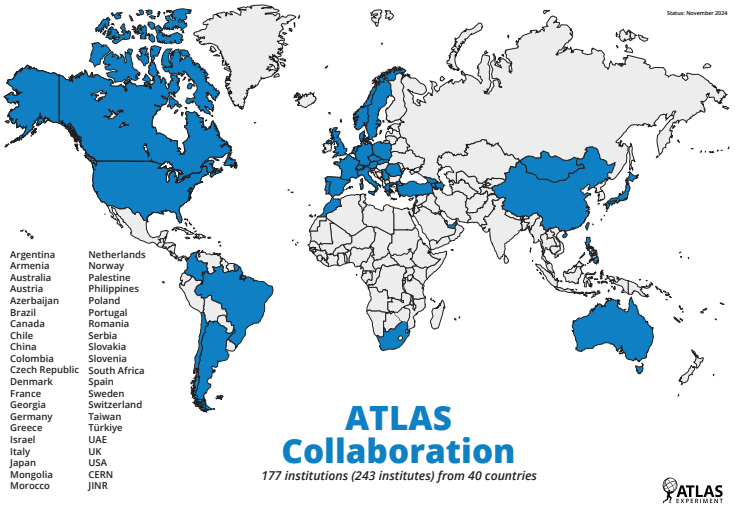
CERN's LHC — general-purpose detector ATLAS



ATLAS detector captures
details of particle interactions
in proton-proton collisions

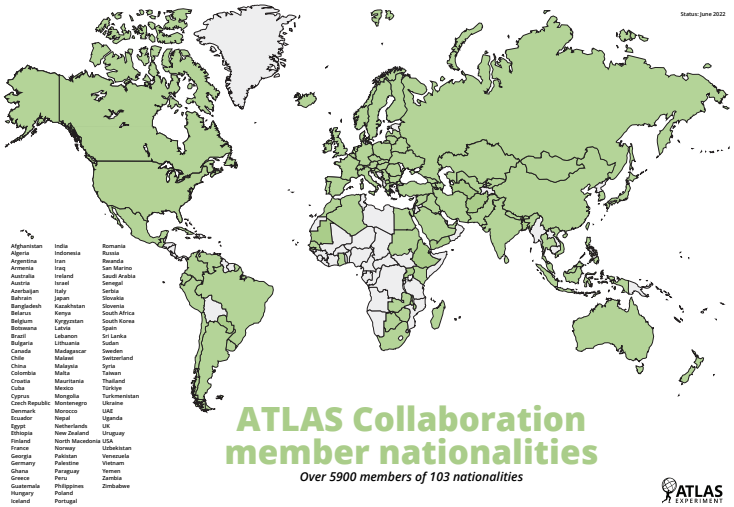
- cylindrical, 44 m long, 25 m in diameter, weighing 7000 t (like Eiffel Tower)
- proton beams cross each other every 25 ns, single event has millions of detector channels
- triggering reduces data rate from 40 MHz (1 PB s^{-1}) to 200 Hz (320 MB s^{-1})
- each experiment stores 10 PB of data per year

ATLAS Collaboration



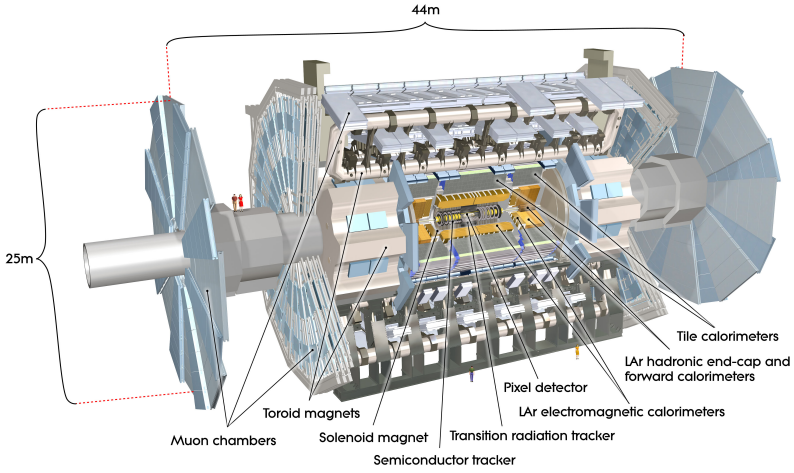
ATLAS Collaboration
involves 6003 members
(3822 physicists)
from 243 institutes
in 40 countries

ATLAS Collaboration — member nationalities



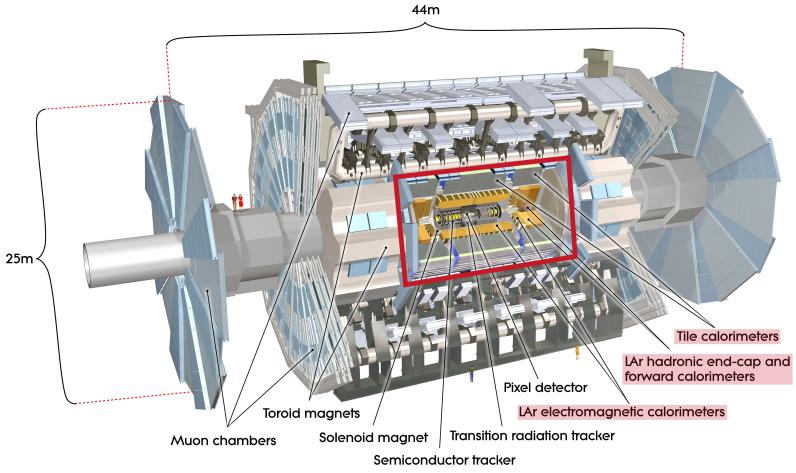
ATLAS Collaboration
involves 6003 members
(3822 physicists)
from 243 institutes
in 40 countries

ATLAS — a toroidal LHC apparatus



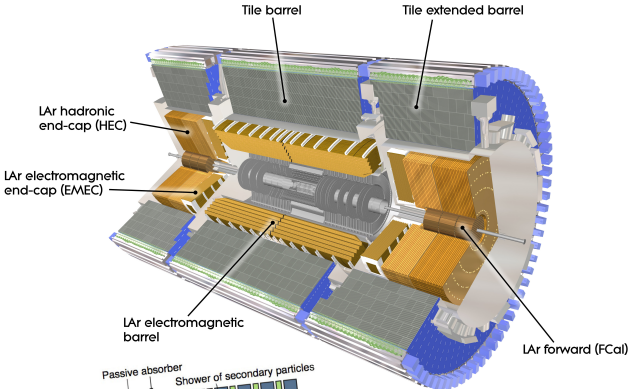
six different detecting subsystems wrapped concentrically in layers around collision point to record trajectory, momentum, and energy of particles

ATLAS — a toroidal LHC apparatus



inner tracking detector, electromagnetic and hadronic calorimeters (ECAL/HCAL), and muon spectrometer

ATLAS energy measurements — sampling calorimeters

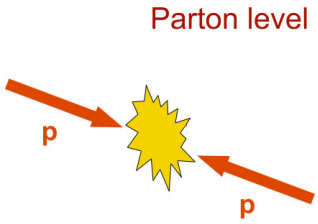


Calorimeters measure energies
by fully absorbing particles in detector

- dense passive absorber stops incoming particles and initiates cascade of secondary particles
- shower can be **electromagnetic** (electrons, positrons, photons) or **hadronic** (quark-bound particles)
- active material produces detectable electric signal from particle shower
- only part of total deposited energy is “sampled”

Calorimeter-signal calibration (topo-clusters)

What happens in a proton-proton collision?

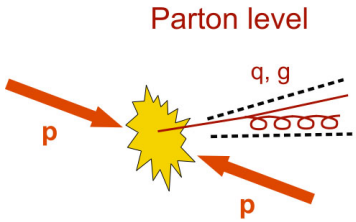


hard-scattering process → fragmentation, parton showering → hadronisation, particle jets → detector response (topo-clusters)

↑

interaction of high-energy partons (quarks and gluons) inside the colliding protons

What happens in a proton-proton collision?

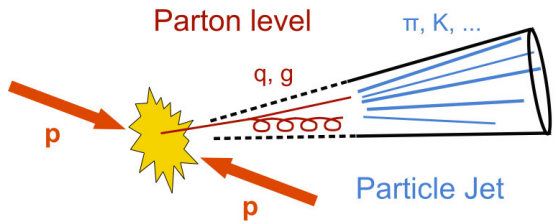


hard-scattering process → **fragmentation, parton showering** → hadronisation, particle jets → detector response (topo-clusters)

↑

high-energy partons **radiate additional partons** and radiated partons further produce more partons, leading to a **cascade of lower-energy partons** (parton shower)

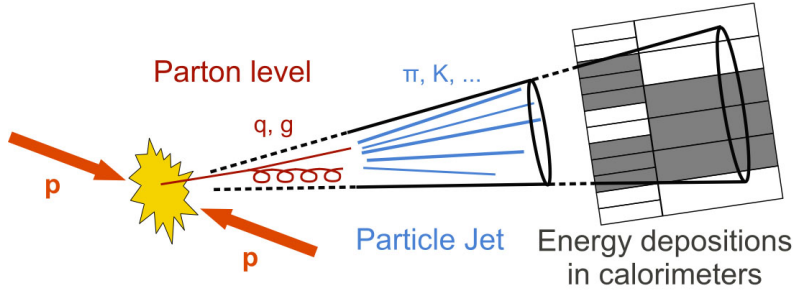
What happens in a proton-proton collision?



hard-scattering process → fragmentation, parton showering → **hadronisation, particle jets** → detector response (topo-clusters)

transition from invisible color-charged partons to **observable color-neutral hadrons** (baryons and mesons), followed by successive hadron decays

What happens in a proton-proton collision?

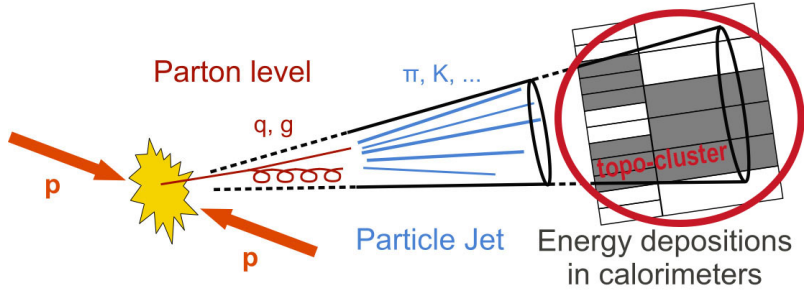


hard-scattering process → fragmentation, parton showering → hadronisation, particle jets → **detector response (topo-clusters)**

↑

hadrons enter detector and deposit energy in calorimeters (e.g. protons, neutrons, pions)

Topo-clusters — topologically connected calorimeter cells

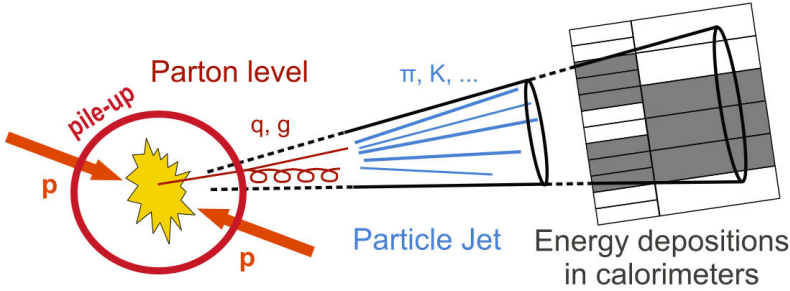


hard-scattering process → fragmentation, parton showering → hadronisation, particle jets → **detector response (topo-clusters)**

individual energy depositions of collision event are grouped into three-dimensional **clusters of topologically connected calorimeter-cell signals** (energy blobs)

using nearest-neighbour algorithms following spatial signal-significance patterns

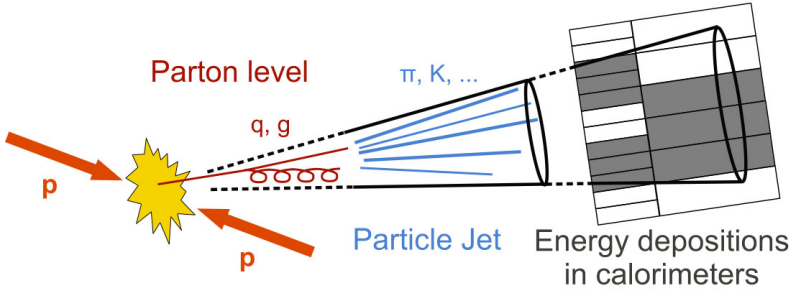
Pile-up interactions — unwanted extra collisions



hard-scattering process → fragmentation, parton showering → hadronisation, particle jets → detector response (topo-clusters)

calorimeter-signal contributions from **additional proton-proton interactions** within same (in-time) or neighbouring/nearby (out-of-time) bunch crossings

Energy calibration for calorimeter signals



$$E_{\text{clus}}^{\text{dep}} \xleftarrow{\text{local energy calibration}} E_{\text{clus}}^{\text{EM}}$$

energy deposited inside cluster's calorimeter cells

basic cluster signal at electromagnetic (EM) scale

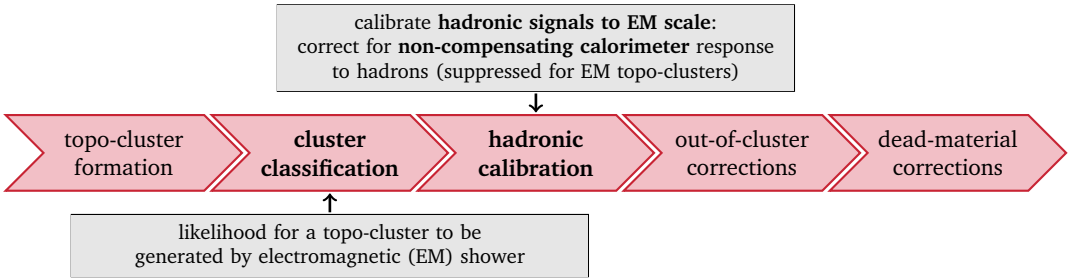
rescale measured calorimeter signals to obtain total deposited energy
(correcting for signal inefficiencies)

Sources of signal inefficiency



non-compensation	**calorimeter signal for hadrons is smaller** than signal for electrons, positrons, and photons depositing same energy (no compensation for **invisible energy losses** from nuclear break-up or nuclear excitation in complex hadronic showers)
pile-up interactions	calorimeter-signal contributions (energy deposits) from **additional proton-proton interactions** in collision environment
clustering algorithm	signal losses due to clustering strategy, in particular at cluster margins (energy lost due to **cells missed in cluster formation**, “lost cells”)
dead material	signal losses due to **energy deposited in inactive, passive or dead material** in front, between, and inside the calorimeter modules

LCW — local cell-weighting calibration for topo-clusters



LCW is four-step sequence with multi-dimensional binned look-up tables

- non-smooth step-like transitions between scale factors,
  no feature correlations (average scale factors in bins), no pile-up measures
- energy calibration depends on nature of cluster-generating shower (EM or hadronic)
  → EM: compact shower development, early starting point, small intrinsic fluctuations
  → shower depth λ_{clus} and cell-signal density $\langle \rho_{\text{cell}} \rangle$ most efficient for classification

Bayesian neural networks

Topo-cluster calibration with neural networks



Why topo-cluster calibration?

[arXiv:1603.02934, ATL-PHYS-PUB-2023-019]

- calorimeter signals calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers
- deposited energy from measured signals

$$E_{\text{clus}}^{\text{EM}} \xrightarrow{\text{hadronic calibration}} E_{\text{clus}}^{\text{had}} = E_{\text{clus}}^{\text{dep}}$$

multi-dimensional correlated calibration:
regression network learns response
over topo-cluster phase space

$$\mathcal{R}_{\text{clus}}^{\text{BNN}}(\mathcal{X}_{\text{clus}}) \overset{\text{train}}{\approx} \mathcal{R}_{\text{clus}}^{\text{EM}} = \frac{E_{\text{clus}}^{\text{EM}}}{E_{\text{clus}}^{\text{dep}}}$$

Topo-cluster calibration with neural networks

Why topo-cluster calibration?

[arXiv:1603.02934, ATL-PHYS-PUB-2023-019]

- calorimeter signals calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers
- deposited energy from measured signals

$$E^{\text{EM}}_{\text{clus}} \xrightarrow{\text{hadronic calibration}} E^{\text{had}}_{\text{clus}} = E^{\text{dep}}_{\text{clus}}$$

multi-dimensional correlated calibration:
regression network learns response
over topo-cluster phase space

$$\mathcal{R}^{\text{BNN}}_{\text{clus}}(\mathcal{X}_{\text{clus}}) \overset{\text{train}}{\approx} \mathcal{R}^{\text{EM}}_{\text{clus}} = \frac{E^{\text{EM}}_{\text{clus}}}{E^{\text{dep}}_{\text{clus}}}$$

15 physics-motivated cluster features → dataset D_{train} given by pairs $(\mathcal{X}_{\text{clus}}, \mathcal{R}^{\text{EM}}_{\text{clus}})$

$$\mathcal{X}_{\text{clus}} = \Big\{ \underbrace{E^{\text{EM}}_{\text{clus}}, y^{\text{EM}}_{\text{clus}}}_{\text{kinematics}}, \overbrace{\zeta^{\text{EM}}_{\text{clus}}}^{\text{signal relevance}}, \underbrace{\text{Var}_{\text{clus}}(t_{\text{cell}}), \lambda_{\text{clus}}, |\vec{c}_{\text{clus}}|, \langle\rho_{\text{cell}}\rangle, \langle \text{m}^2_{\text{long}}\rangle, \langle \text{m}^2_{\text{lat}}\rangle, p_{\text{T}}D, f_{\text{emc}}, \overbrace{f_{\text{iso}}}^{\text{topology (isolation)}}}_{\text{shower nature (position, compactness, signal density, internal time structure)}}, \underbrace{t_{\text{clus}}, N_{\text{PV}}, \mu}_{\text{pile-up}} \Big\}$$

Topo-cluster calibration with neural networks

Why topo-cluster calibration?

[arXiv:1603.02934, ATL-PHYS-PUB-2023-019]

- calorimeter signals calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers
- deposited energy from measured signals

$$E^{\text{EM}}_{\text{clus}} \xrightarrow{\text{hadronic calibration}} E^{\text{had}}_{\text{clus}} = E^{\text{dep}}_{\text{clus}}$$

Modern **(B)NNs** for
local hadronic topo-cluster calibration
correcting for this non-compensation

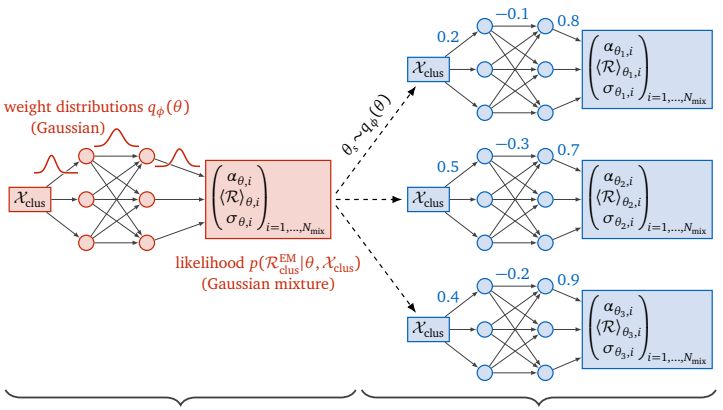
- single-step training
- exploiting correlations
- smooth and multi-dimensional
- **control and uncertainties key**
  (access to bottom-up systematics)

[Ph.D. Thesis of Y. Gal, arXiv:2211.01421]

learn uncertainty associated with calibration function

topo-cluster selections; error propagation; better understanding of signal-quality issues in weak detector regions, and of network-training limitations

BNN — Bayesian neural network



BNNs **learn distributions** of network parameters, defining output distribution

[arXiv:2003.11099, arXiv:2206.14831, arXiv:2211.01421]

- weights are not trained as fixed values
- training: parameters θ described by weight distributions $q_\phi(\theta) \approx p(\theta|D_{\text{train}})$
- inference: sample from weight distributions to get ensemble of networks

BNN — learning weight distributions



BNNs learn distributions of network parameters, defining output distribution

$\mathcal{R}^{\text{BNN}}_{\text{clus}}(\mathcal{X}_{\text{clus}})$ given by probability $p(\mathcal{R})$ encoded in weight posterior:

$$p(\mathcal{R}) = \int \mathrm{d}\theta \; p(\mathcal{R}|\theta)\, p(\theta|D_{\text{train}})$$

BNN — variational inference



BNNs learn distributions of network parameters, defining output distribution

$\mathcal{R}^{\text{BNN}}_{\text{clus}}(\mathcal{X}_{\text{clus}})$ given by probability $p(\mathcal{R})$ encoded in weight posterior:

$$p(\mathcal{R}) = \int \mathrm{d}\theta \; p(\mathcal{R}|\theta) \; p(\theta|D_{\text{train}}) \approx \int \mathrm{d}\theta \; p(\mathcal{R}|\theta) \; q_\phi(\theta)$$

but weight posterior is intractable...
training by **variational approximation** of
$p(\theta|D_{\text{train}})$ with simplified and tractable $q_\phi(\theta)$

BNN — loss function



BNNs learn distributions of network parameters, defining output distribution

$\mathcal{R}^{\text{BNN}}_{\text{clus}}(\mathcal{X}_{\text{clus}})$ given by probability $p(\mathcal{R})$ encoded in weight posterior:

$$p(\mathcal{R}) = \int \mathrm{d}\theta \; p(\mathcal{R}|\theta)\, p(\theta|D_{\text{train}}) \approx \int \mathrm{d}\theta \; p(\mathcal{R}|\theta)\, q_\phi(\theta)$$

Similarity by minimizing reverse **Kullback-Leibler (KL) divergence**:

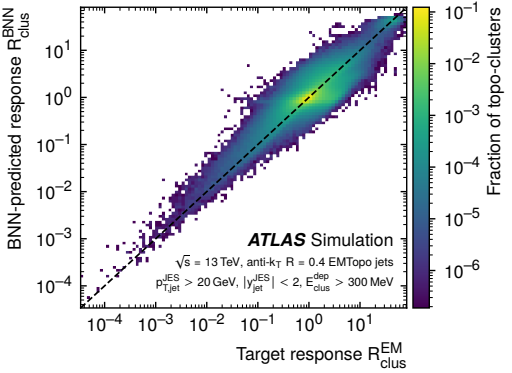
$$\min_\phi D_{\text{KL}}[q_\phi(\theta), p(\theta|D_{\text{train}})] \xrightarrow{\text{Bayes}} \mathcal{L}_{\text{BNN}} = \underbrace{D_{\text{KL}}[q_\phi(\theta), p_{\text{prior}}(\theta)]}_{\text{weight regularization}} - \underbrace{\left\langle \log p(D_{\text{train}}|\theta) \right\rangle_{\theta\sim q_\phi(\theta)}}_{\text{negative log-likelihood}}$$

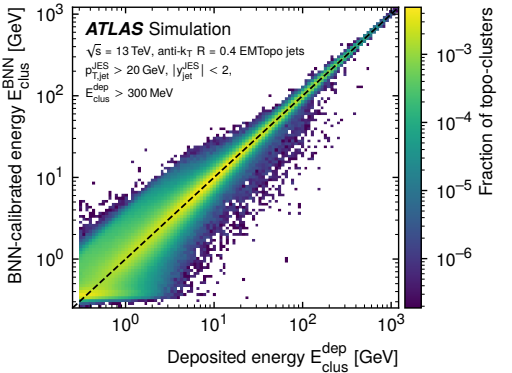
BNN-calibration performance

BNN — response prediction and energy calibration

$$\mathcal{R}^{\text{BNN}}_{\text{clus}} \xrightarrow{\text{train}} \mathcal{R}^{\text{EM}}_{\text{clus}} = \frac{E^{\text{EM}}_{\text{clus}}}{E^{\text{dep}}_{\text{clus}}}$$

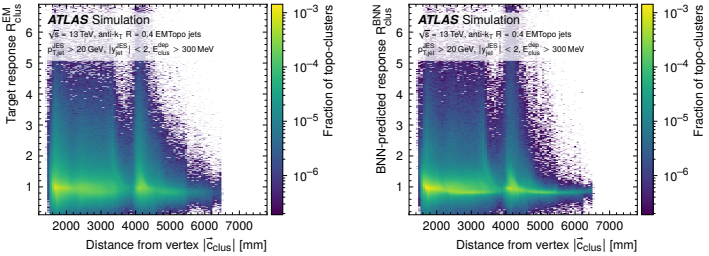
$$E^{\text{BNN}}_{\text{clus}} = \frac{E^{\text{EM}}_{\text{clus}}}{\mathcal{R}^{\text{BNN}}_{\text{clus}}} \longrightarrow E^{\text{dep}}_{\text{clus}}$$





agreement of BNN prediction and regression target:
correlation curves for predicted response and calibrated energy look promising

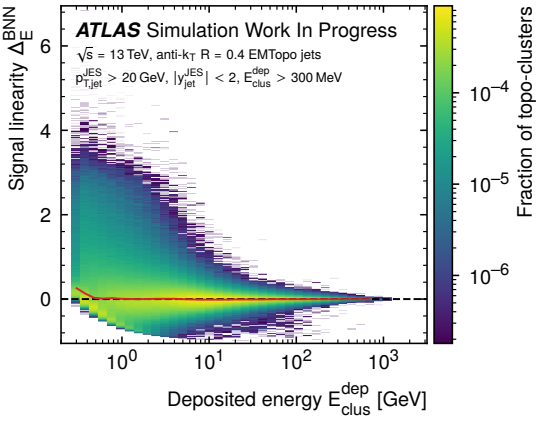
BNN — response vs features



very complex target distributions (i.e. due to calorimeter geometry) are reproduced well by the BNN

example: transition from central barrel to endcap calorimeter in $|\vec{c}_{\text{clus}}|$

BNN — signal linearity

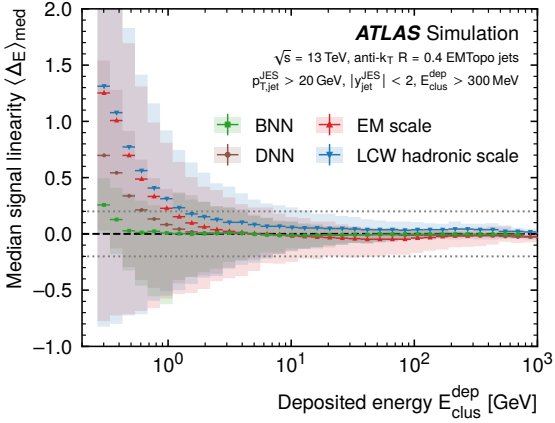


Signal linearity: ratio of calibrated over deposited energy

$$\Delta_E^\kappa = \frac{E_{\text{clus}}^\kappa}{E_{\text{clus}}^{\text{dep}}} - 1 \quad \text{with} \quad E_{\text{clus}}^\kappa = \frac{E_{\text{clus}}^{\text{EM}}}{\mathcal{R}_{\text{clus}}^\kappa}$$

- scales $\kappa \in \{\text{EM}, \text{LCW}, \text{DNN}, \text{BNN}\}$
- should peak at zero after successful calibration
- evaluated as function of features and deposited energy
- measure for quality or accuracy of learned calibration

BNN — median signal linearity

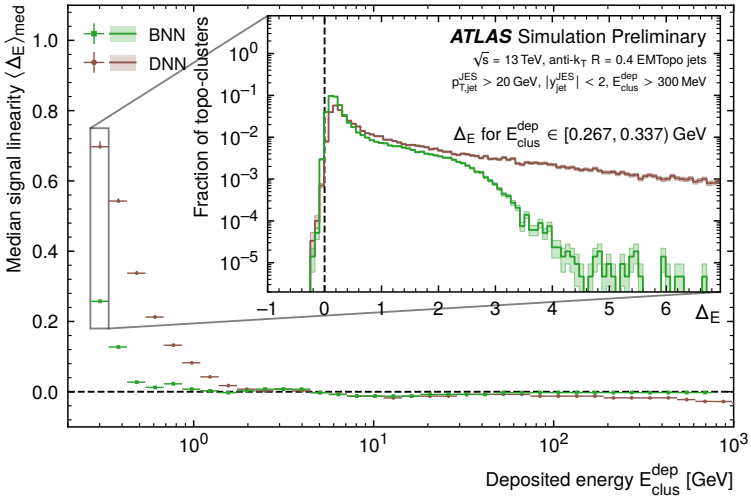


Signal linearity: ratio of calibrated over deposited energy

$$\Delta_E^\kappa = \frac{E_{\text{clus}}^\kappa}{E_{\text{clus}}^{\text{dep}}} - 1 \quad \text{with} \quad E_{\text{clus}}^\kappa = \frac{E_{\text{clus}}^{\text{EM}}}{\mathcal{R}_{\text{clus}}^\kappa}$$

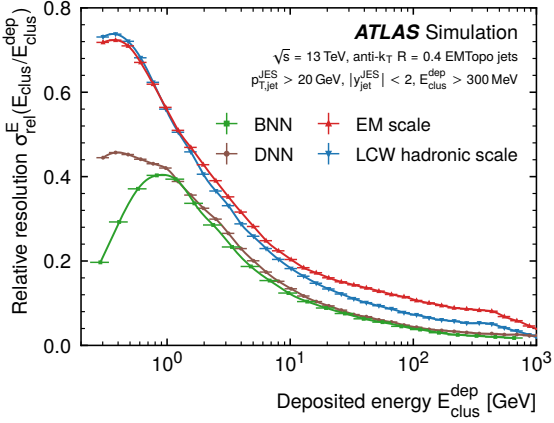
- scales $\kappa \in \{\text{EM}, \text{LCW}, \text{DNN}, \text{BNN}\}$
- should peak at zero after successful calibration
- evaluated as function of features and deposited energy
- BNN better over whole energy range, most significant at low energies

BNN — bin-wise signal linearity



BNN-derived calibration shows significantly suppressed tails compared to DNN

BNN — relative local energy resolution

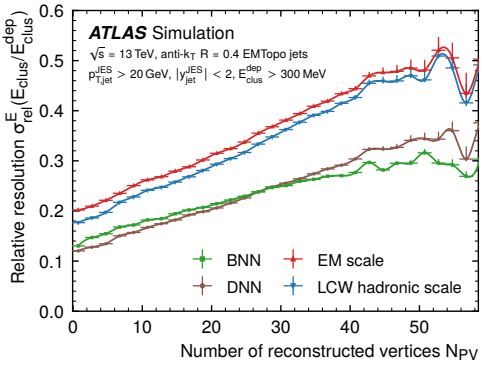


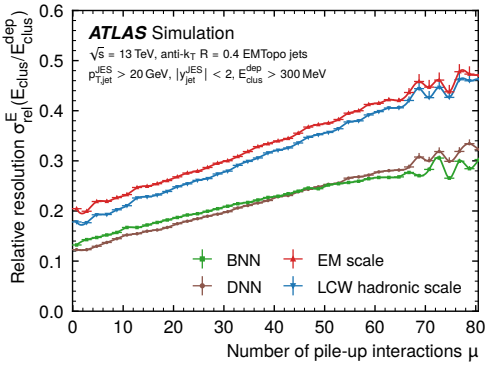
Relative local energy resolution:

$$\sigma^{E}_{\text{rel}} = \frac{Q^{w}_{f=68\,\%}}{2\langle E^{\kappa}_{\text{clus}}/E^{\text{dep}}_{\text{clus}}\rangle_{\text{med}}}$$

- $Q^{w}_{f=68\,\%} \equiv 68\,\%$ inter-quantile range
- BNN better over whole energy range, spectacular at low energies
- BNN learns signal-source transition from inelastic hadronic interactions to ionisation-dominated signals

BNN — relative local energy resolution

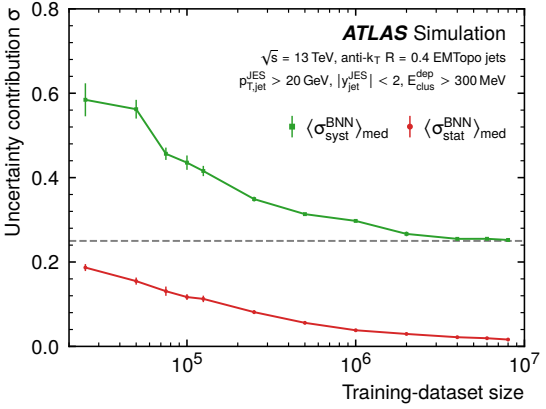




relative local energy resolution vs in-time and out-of-time pile-up activity
→ ML-derived calibration shows cluster-by-cluster pile-up mitigation

BNN-learned uncertainties

BNN — systematic and statistical uncertainties



Learned σ_{tot} with two origins:

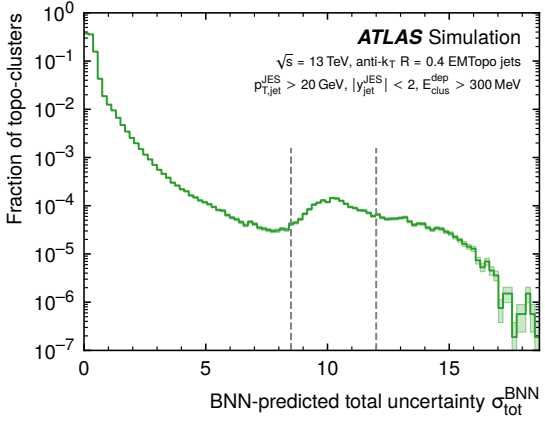
[arXiv:1904.10004, arXiv:2003.11099, arXiv:2206.14831]

- **statistics σ_{stat}** $\rightarrow p(\theta|D_{\text{train}})$
  limited training statistics (epistemic),
  vanishing for infinite training data
- **systematics σ_{syst}** $\rightarrow p(\mathcal{R}^{\text{EM}}_{\text{clus}}|\theta, \mathcal{X}_{\text{clus}})$
  intrinsic randomness from stochastic training data (pile-up; aleatoric),
  limited network expressivity,
  non-optimal hyper-parameters,
  plateau for good training statistics

For well-trained LHC networks:

$$\sigma_{\text{tot}} \equiv \sqrt{\sigma^2_{\text{syst}} + \sigma^2_{\text{stat}}} \approx \sigma_{\text{syst}} \gg \sigma_{\text{stat}}$$

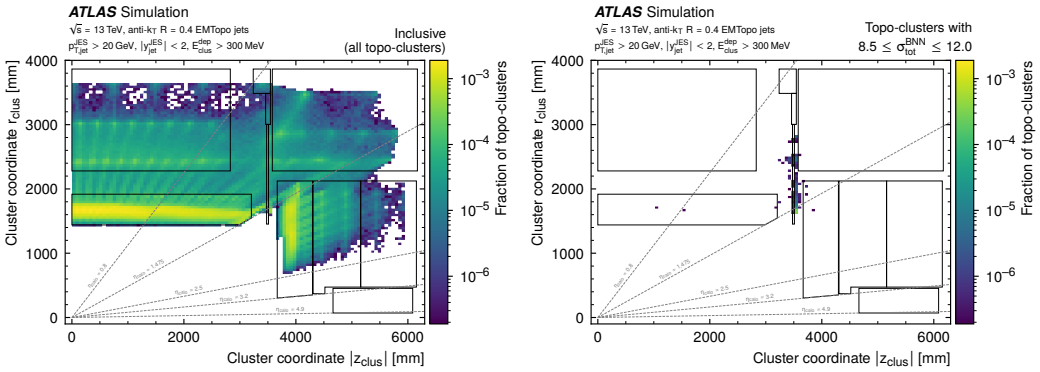
BNN — uncertainties as explainable AI



Use BNN uncertainty to understand data

- uncertainty spectrum shows distinctive secondary maximum, deviating from expected smoothly falling spectrum...
- which feature leads the BNN to flag these topo-clusters with large learned uncertainties?
- analyze anomalous clusters in terms of detector geometry

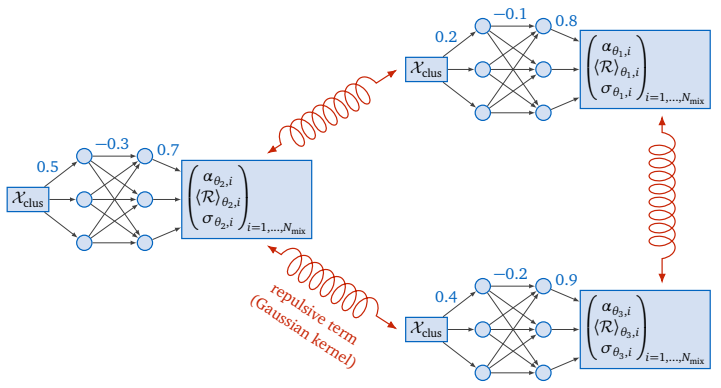
BNN — uncertainties as explainable AI



large uncertainties from complex transition region in calorimeter system:
tile-gap scintillator not a regular calorimeter → feature quality in this region is insufficient

Repulsive ensembles

RE — repulsive ensemble



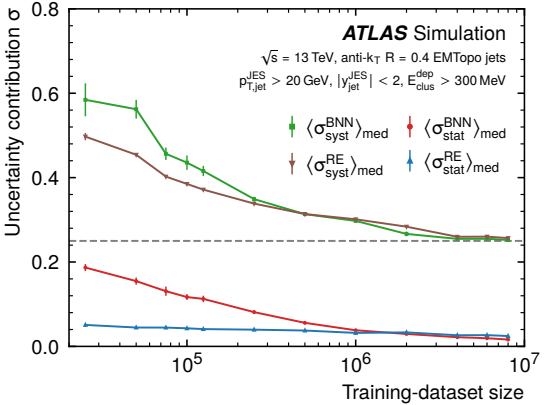
training: repulsive term connecting function space of all simultaneously trained networks forces ensemble to spread out and cover loss around actual minimum

Alternative way
for uncertainty estimation

[arXiv:2106.11642, arXiv:2211.01421, arXiv:240313899]

- regular ensembles do not sample from weight posterior
- introduce repulsive force between ensemble members during optimization such that $\theta \sim p(\theta|D_{\text{train}})$
- **repulsive term** ensures that uncertainty covers probability distribution over space of network functions

RE — systematic and statistical uncertainties

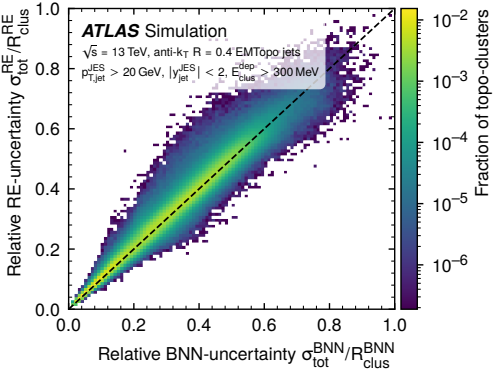


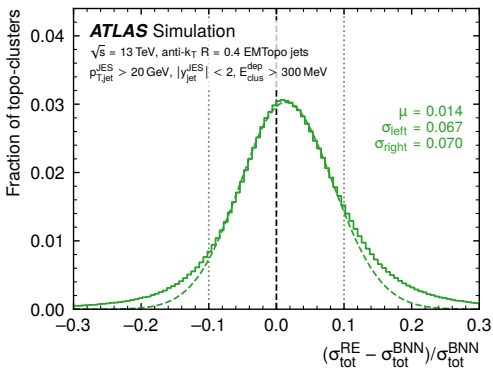
Repulsive force ensures that uncertainty covers probability distribution over space of network functions

[arXiv:2106.11642, arXiv:2211.01421, arXiv:240313899]

- gives two uncertainties
- **systematics σ_{syst}**
  plateau for good training statistics, part of likelihood (same as for BNN)
- **statistics σ_{stat}**
  vanishing for good training statistics (but with flatter slope)

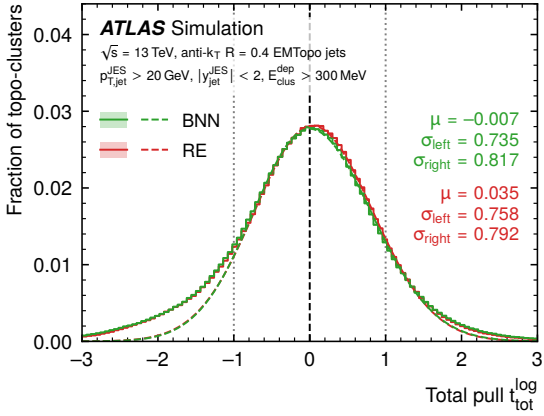
BNN and RE — consistency check





10 % agreement between ML-based uncertainty estimates
and both uncertainty predictions track each other well

BNN and RE — uncertainties vs data spread



Pull: central prediction and uncertainty
Does uncertainty cover data spread?

$$t^{\kappa}_{\text{tot}}(x) = \frac{\mathcal{R}^{\kappa}_{\text{clus}}(x) - \mathcal{R}^{\text{EM}}_{\text{clus}}(x)}{\sigma^{\kappa}_{\text{tot}}(x)}$$

- evaluated in $\log_{10} \mathcal{R}_{\text{clus}}$ space
- stochastic data defining shape, Gaussian with order-one width
- BNN and RE errors consistent
- per-cluster error conservative (reliable upper limit for deviation)

Summary and outlook

Summary and outlook



EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: SciPost Phys.

CERN-EP-2024-317
6th December 2024

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

The ATLAS Collaboration[1]

The ATLAS experiment at the Large Hadron Collider explores the use of modern neural networks for a multi-dimensional calibration of its calorimeter signal defined by clusters of topologically connected cells (topo-clusters). The Bayesian neural network (BNN) approach not only yields a continuous and smooth calibration function that improves performance relative to the standard calibration but also provides uncertainties on the calibrated energies for each topo-cluster. The results obtained by using a trained BNN are compared to the standard local hadronic calibration and to a calibration provided by training a deep neural network. The uncertainties predicted by the BNN are interpreted in the context of a fractional contribution to the systematic uncertainties of the trained calibration. They are also compared to uncertainty predictions obtained from an alternative estimator employing repulsive ensembles.

© 2024 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

[1] The full author list can be found at:
https://atlas.web.cern.ch/Atlas/PUBNOTES/ATL-PHYS-PUB-2024-XXX/authorlist.pdf

arXiv:2412.04370v1 [hep-ex] 5 Dec 2024

Modern uncertainty-aware BNNs for multi-dimensional calorimeter-signal calibration

- continuous and smooth calibration of topo-clusters
- improved performance relative to LCW and DNN
- meaningful per-cluster systematics
- BNN and RE learn reliable uncertainties

Next steps:

- try deterministic NN with heteroscedastic loss instead of BNN (to speed up inference) [arXiv:2412.12069]
- further tune (B)NN performance
- full performance study within ATLAS (apply simulation-trained calibration to data)

References and further reading

ML-based topo-cluster calibration and ML with uncertainties

- ATLAS Collaboration
  The application of neural networks for the calibration of topological cell clusters in the ATLAS calorimeters
  ATLAS PUB Note (2023)
- ATLAS Collaboration (including T. Heimel, P. Loch, T. Plehn and L. Vogel)
  Precision calibration of calorimeter signals in the ATLAS experiment using an uncertainty-aware neural network
  arXiv:2412.04370 [hep-ex]
- Y. Gal
  Uncertainty in Deep Learning
  Ph.D. Thesis, University of Cambridge (2016)
- T. Plehn, A. Butter, B. Dillon, T. Heimel, C. Krause and R. Winterhalder
  Modern Machine Learning for LHC Physicists
  arXiv:2211.01421 [hep-ph] (continuously updated on website)

References and further reading

Bayesian neural networks (BNNs) and repulsive ensembles (REs)

- G. Kasieczka, M. Luchmann, F. Otterpohl and T. Plehn
  Per-object systematics using deep-learned calibration
  SciPost Phys. 9, 089 (2020), arXiv:2003.11099 [hep-ph]
- H. Bahl, N. Elmer, L. Favaro, M. Haußmann, T. Plehn and R. Winterhalder
  Accurate Surrogate Amplitudes with Calibrated Uncertainties
  arXiv:2412.12069 [hep-ph]
- F. D'Angelo and V. Fortuin
  Repulsive Deep Ensembles are Bayesian
  arXiv:2106.11642 [cs.LG]
- L. Röver, B. M. Schäfer and T. Plehn
  PINNferring the Hubble Function with Uncertainties
  arXiv:2403.13899 [astro-ph.CO]

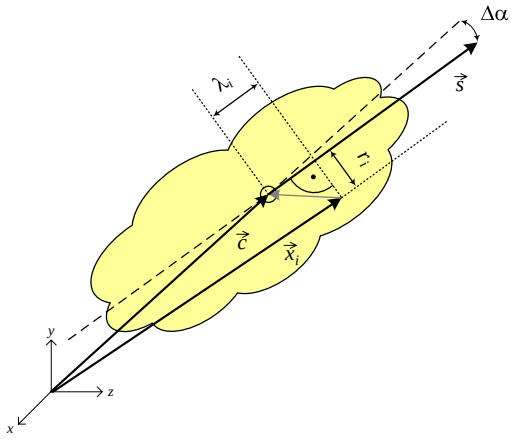
Backup slides...

Dataset — topo-cluster features

Table 1: Dataset consists of topo-clusters reconstructed in Monte Carlo (MC) simulations of full proton-proton collision events at $\sqrt{s} = 13\,\text{TeV}$ (LHC Run 2 pile-up) with multi-jet final states

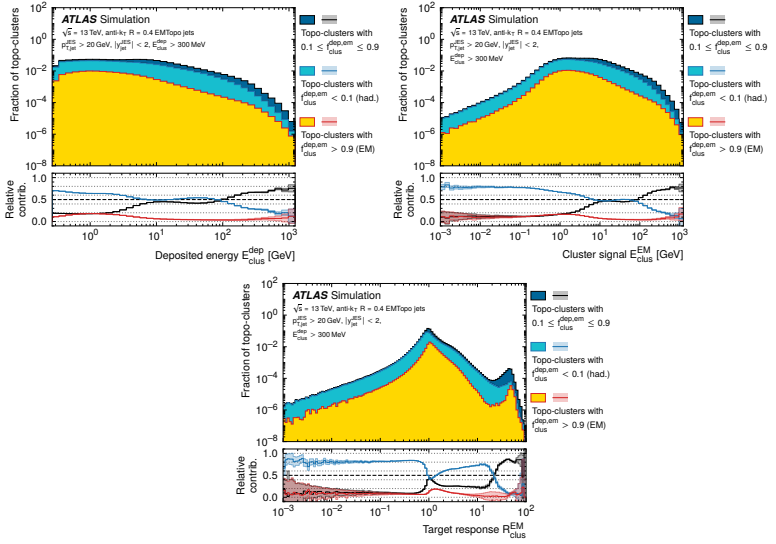
category	symbol	description / comment
kinematics	$E^{\text{EM}}_{\text{clus}}$, $y^{\text{EM}}_{\text{clus}}$	cluster signal and rapidity at electromagnetic (EM) energy scale
signal strength	$\zeta^{\text{EM}}_{\text{clus}}$	signal significance
timing	t_{clus}	cluster-signal time
time structure	$\text{Var}_{\text{clus}}(t_{\text{cell}})$	variance of cell-time distribution in cluster
shower depth	$\lvert\vec{c}_{\text{clus}}\rvert$	distance of centre-of-gravity from nominal vertex
	λ_{clus}	distance of centre-of-gravity from calorimeter front face
shower shape, compactness	f_{emc}	fraction of energy in electromagnetic calorimeter
	$\langle\rho_{\text{cell}}\rangle$, $p_{\text{T}}D$	cluster signal density and signal compactness
	$\langle \text{m}^2_{\text{long}}\rangle$, $\langle \text{m}^2_{\text{lat}}\rangle$	energy dispersion along and perpendicular to main cluster axis
topology	f_{iso}	cluster isolation measure (effects of neighboring/nearby signals)
pile-up	N_{PV}	number of reconstructed primary vertices (in-time pile-up)
	μ	number of interactions per bunch crossing (out-of-time pile-up)

Geometrical moments for topo-clusters

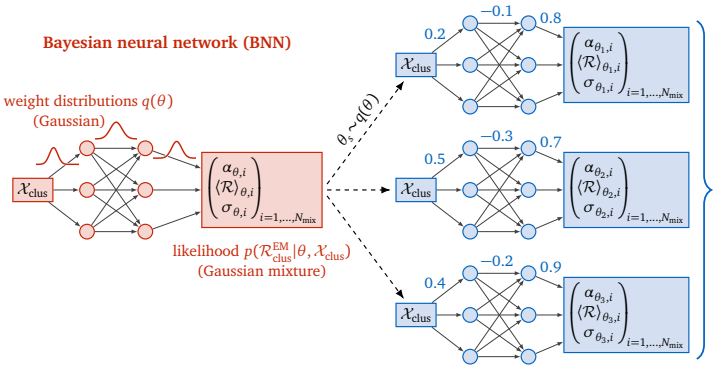


- $\vec{c}$ *centre of gravity of cluster, measured from the nominal vertex* $(x = 0, y = 0, z = 0)$ *in ATLAS*
- $\vec{x}_i$ *geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*
- $\vec{s}$ *particle direction of flight (shower axis)*
- $\Delta\alpha$ *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* $\vec{s}$ $(r_i \geq 0)$

Dataset — energy and response distributions



BNN — network architecture



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

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

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

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

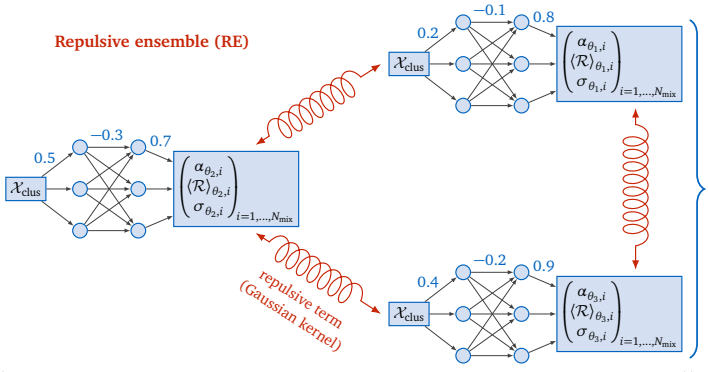
with

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

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

RE — network architecture



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

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

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

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

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: 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

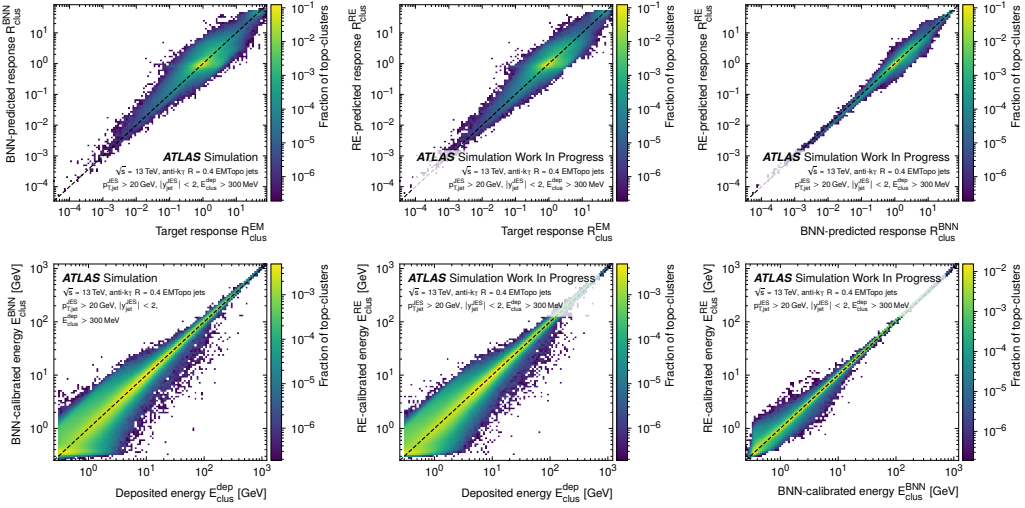
Inference: Same formulas as for the BNN, using the N simultaneously trained ensemble members

BNN — network setup and hyper-parameters

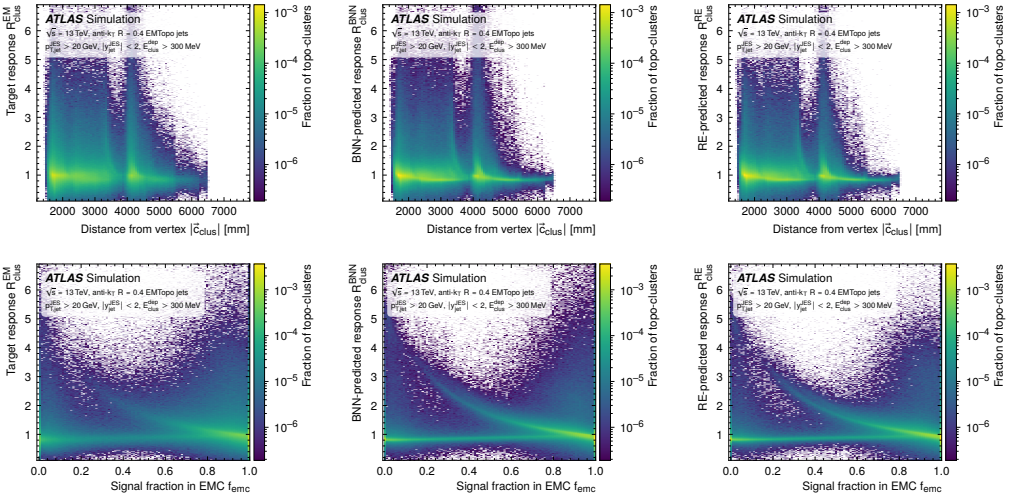
Table 2: BNN and RE setup for three-mode Gaussian mixture likelihood

hyper-parameter	**network architecture and setup**
likelihood model	Gaussian mixture model (GMM)
number of mixture modes N_{mix}	3 (i.e. 9 output nodes)
number of hidden layers	4 (with 64 nodes each)
nodes per layer (input, hidden, and output)	$\{15, 64, 64, 64, 64, 9\}$
inner activation functions	rectified linear unit (ReLU)
central-value prediction	maximum likelihood ("mode")
optimizer and learning rate (LR)	ADAM with LR $= 10^{-4}$
learning-rate scheduler	STEPLR, epochs $\{25, 100\}$, $\gamma = 0.1$
number of training epochs	150
training (and inference) batch size	4096 (512)
dataset sizes for training, validation, testing	{8.7M, 500k, 5.3M}
Monte-Carlo samples at inference N	50

BNN and RE — response and energy correlation curves



EM, BNN and RE — response vs features



DNN, BNN and RE — signal linearity vs features

