# BNNs for Topo-Cluster Calibration

## What theorists can teach us about calorimeter calibration

**Lorenz Vogel**
July 18, 2024

Institute for Theoretical Physics
Heidelberg University

*ATLAS-Heidelberg Meeting, Annweiler/Trifels*

STA project on the application of Bayesian neural networks (BNNs) for the calibration of topological cell clusters in the ATLAS calorimeters

— in collaboration with Theo Heimel, Peter Loch, Tilman Plehn, Jad M. Sardain and Philip Velie (and many thanks to Michel Luchmann for the initial code setup)

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

# BNNs for Topo-Cluster Calibration

## What theorists with very little help from Peter can teach us about calorimeter calibration

**Lorenz Vogel**
July 18, 2024

Institute for Theoretical Physics
Heidelberg University

*ATLAS-Heidelberg Meeting, Annweiler/Trifels*

STA project on the application of Bayesian neural networks (BNNs) for the calibration of topological cell clusters in the ATLAS calorimeters

— in collaboration with Theo Heimel, Peter Loch, Tilman Plehn, Jad M. Sardain and Philip Velie (and many thanks to Michel Luchmann for the initial code setup)

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

# Outline

1. Motivation
2. BNN-calibration performance
3. BNN-learned uncertainties
4. Repulsive ensembles
5. Outlook

# Motivation

# Topo-cluster calibration

**Why topo-cluster calibration?**

- clusters of topologically connected cell signals principal calorimeter signals
- calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers

**multi-dimensional correlated calibration**

$$E_{\text{clus}}^{\text{EM}} \xrightarrow{\text{hadronic calibration}} \underbrace{E_{\text{clus}}^{\text{had}} = E_{\text{clus}}^{\text{dep}}}_{\text{expected/goal}}$$

# Topo-cluster calibration

**Why topo-cluster calibration?**

- clusters of topologically connected cell signals principal calorimeter signals
- calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers

**regression network**
response over phase space

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

15 topo-cluster features $\rightarrow$ dataset $D_{\mathrm{train}}$ given by $(\mathcal{X}_{\mathrm{clus}}, \mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}})$

$$\mathcal{X}_{\mathrm{clus}} = \left\{E^{\mathrm{EM}}_{\mathrm{clus}}, y^{\mathrm{EM}}_{\mathrm{clus}}, \zeta^{\mathrm{EM}}_{\mathrm{clus}}, t_{\mathrm{clus}}, \mathrm{Var}_{\mathrm{clus}}(t_{\mathrm{cell}}), \lambda_{\mathrm{clus}}, |\vec{c}_{\mathrm{clus}}|, f_{\mathrm{emc}}, \langle\rho_{\mathrm{cell}}\rangle, \langle\mathfrak{m}^2_{\mathrm{long}}\rangle, \langle\mathfrak{m}^2_{\mathrm{lat}}\rangle, p_{\mathrm{T}}D, f_{\mathrm{iso}}, N_{\mathrm{PV}}, \mu\right\}$$

# Topo-cluster calibration

**Why topo-cluster calibration?**

- clusters of topologically connected cell signals principal calorimeter signals
- calibrated to correctly measure energy deposited by EM showers
- do not compensate for invisible energy losses in complex hadronic showers

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

- single-step training
- exploiting correlations
- smooth and multi-dimensional
- **control and uncertainties key** [arXiv:2211.01421] (access to bottom-up systematics)

# BNN-calibration performance

# BNNs — Bayesian neural networks

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

$\mathcal{R}(x)$ given by probability $p(\mathcal{R})$ encoded in weight configurations

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

# BNNs — Bayesian neural networks

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

$\mathcal{R}(x)$ given by probability $p(\mathcal{R})$ encoded in weight configurations

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

Training by variational approximation of $p(\theta|D_{\text{train}})$ with simplified and tractable $q(\theta)$

# BNNs — Bayesian neural networks

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

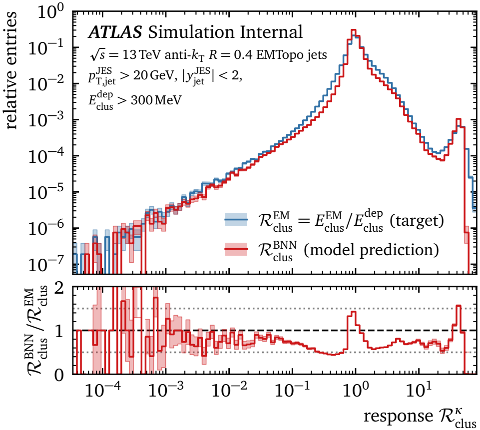
$\mathcal{R}(x)$ given by probability $p(\mathcal{R})$ encoded in weight configurations

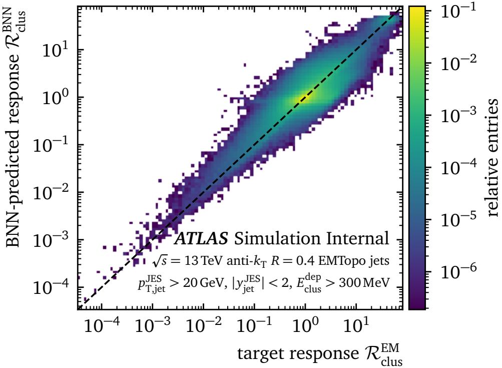
$$p(\mathcal{R}) = \int d\theta \; p(\mathcal{R}|\theta)\, p(\theta|D_{\text{train}}) \approx \int d\theta \; p(\mathcal{R}|\theta)\, q(\theta)$$

Similarity by minimizing KL-divergence

$$\min_{\theta} D_{\text{KL}}[q(\theta), p(\theta|D_{\text{train}})] \xrightarrow{\text{Bayes}} \mathcal{L}_{\text{BNN}} = \underbrace{D_{\text{KL}}[q(\theta), p(\theta)]}_{\text{regularization}} \underbrace{- \Big\langle \log p(D_{\text{train}}|\theta) \Big\rangle_{\theta\sim q(\theta)}}_{\text{log-likelihood}}$$

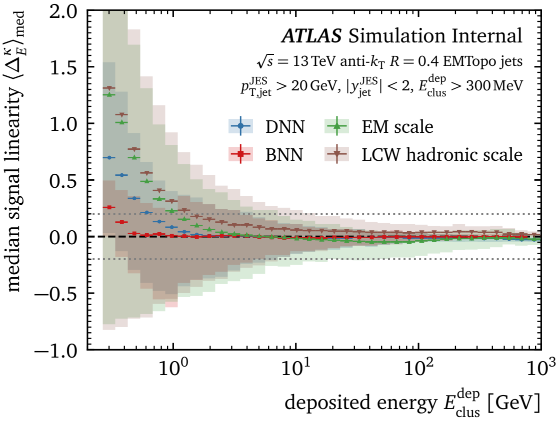
# BNN — response prediction




target-response shape challenging but correlation curve looks promising

# BNN — median signal linearity

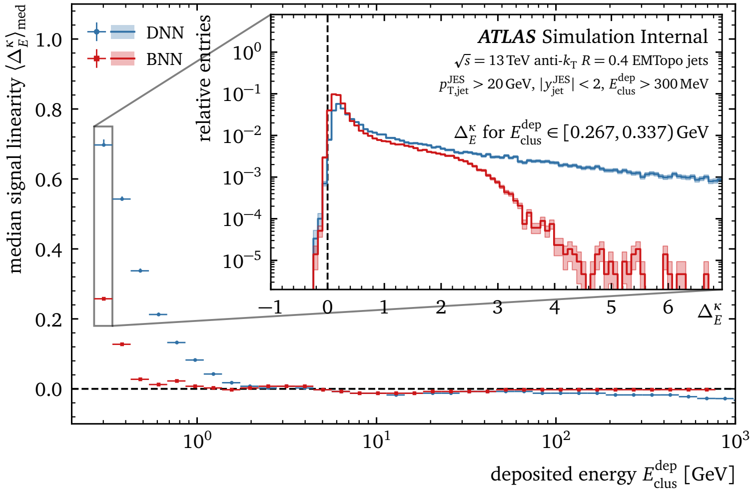



Signal linearity: ratio of calibrated over deposited energy

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

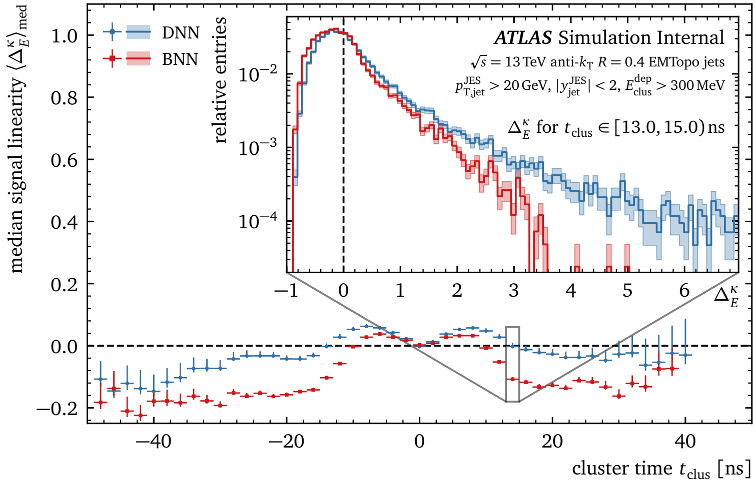
- BNN better over whole energy range
- most significant at low energies

# BNN — bin-wise signal linearity



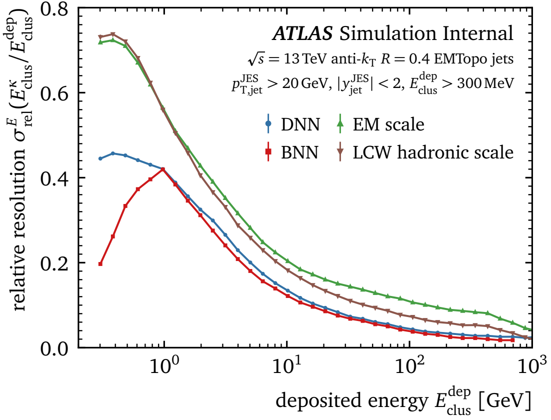

BNN better over whole energy range — most significant at low energies

# BNN — bin-wise signal linearity




summarizing a whole distribution in one parameter is not sufficient...

# BNN — relative local energy resolution

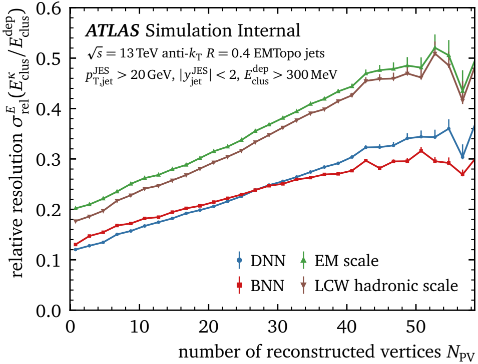


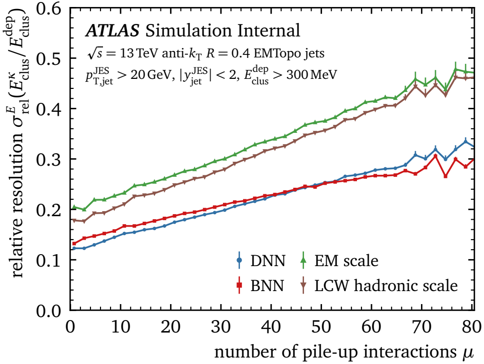

Relative local energy resolution

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

- BNN better over whole energy range
- spectactular at low energies

# BNN — relative local energy resolution

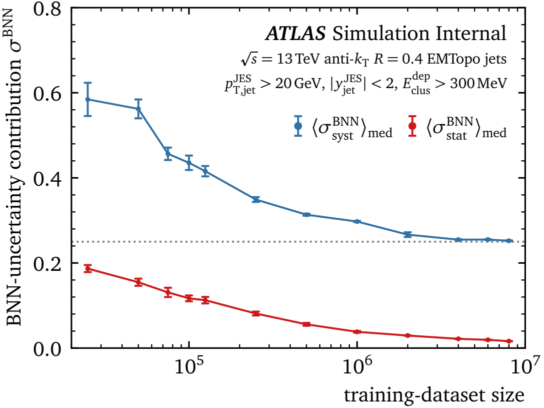



relative local energy resolution vs pile-up → BNN pile-up mitigation?

# BNN-learned uncertainties

# BNN — systematic and statistical uncertainties




Learned $\sigma_{tot}$ with two origins

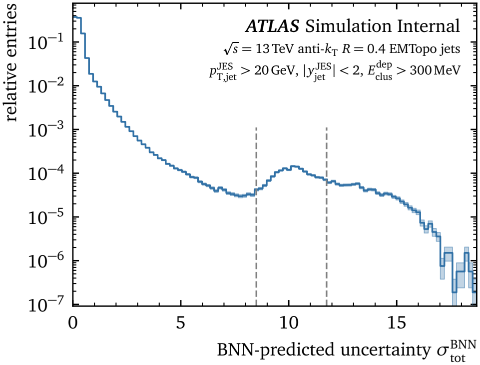
[arXiv:1904.10004, arXiv:2003.11099, arXiv:2206.14831]

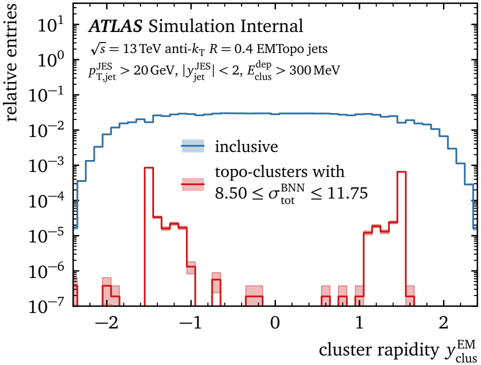
- statistics $\sigma_{stat}$
  training statistics,
  vanishing for good training statistics
- systematics $\sigma_{syst}$
  stochastic training data,
  limited network expressivity,
  bad hyper-parameters,
  plateau for good training statistics

Well-trained LHC models

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

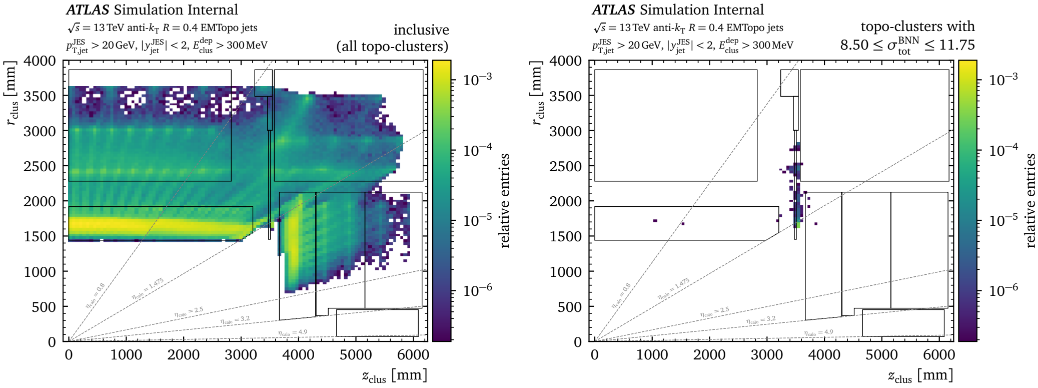
# BNN — uncertainties as xAI




anomalous BNN-uncertainties → use BNN output to understand data...

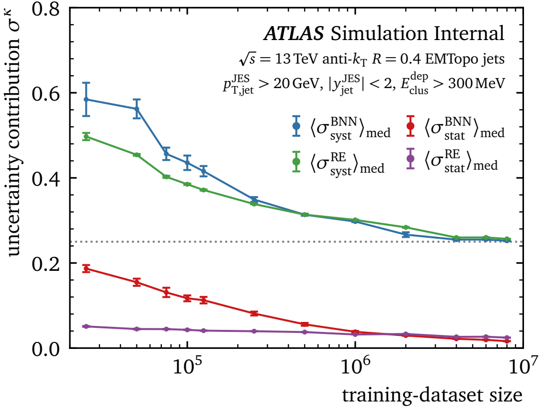
# BNN — uncertainties as xAI




large uncertainties from tile-gap scintillator region

# Repulsive ensembles

# REs — repulsive ensembles

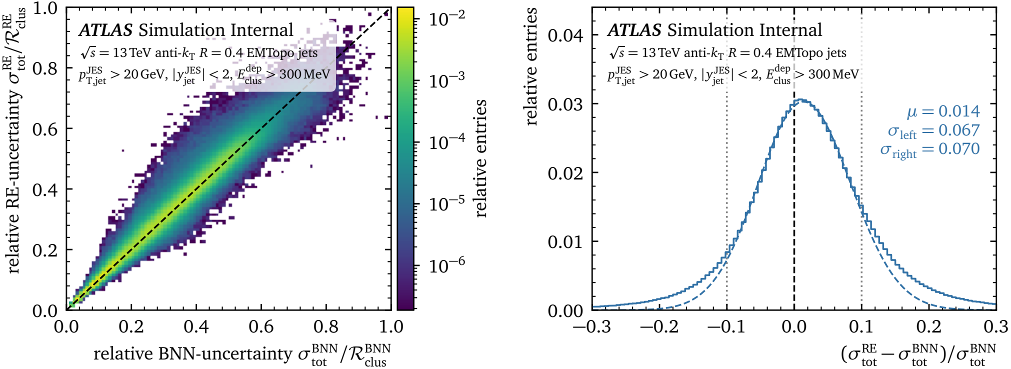



Alternative way for uncertainty estimation

[arXiv:2106.11642, arXiv:2403.13899]

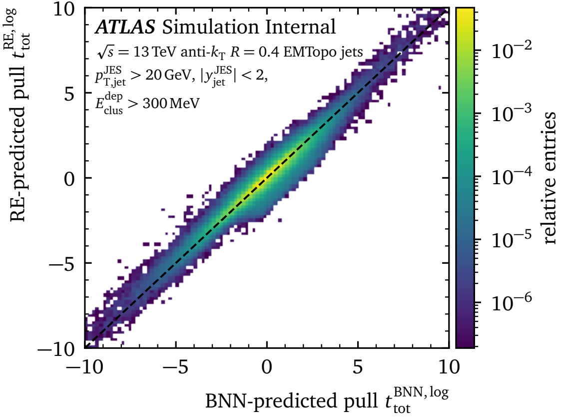
- gives two uncertainties
- systematics $\sigma_{syst}$
  plateau for good training statistics
  part of likelihood (same as for BNN)
- statistics $\sigma_{stat}$
  vanishing for good training statistics
  (with flatter slope)

# BNN and RE — consistency check




10 % agreement between uncertainty estimates

# BNN and RE — uncertainties vs data spread

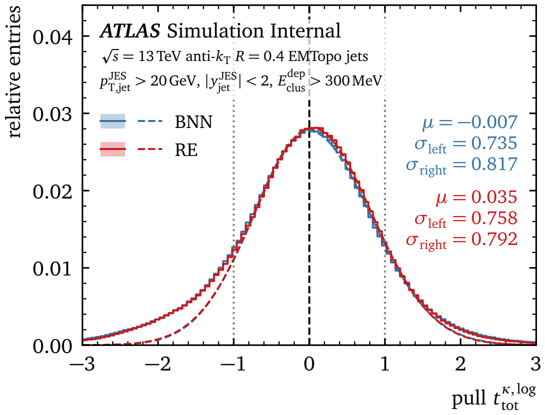



Central prediction plus uncertainty as pull (evaluated in $\log_{10} \mathcal{R}_{\text{clus}}$ space)

$$t_{\text{tot}}^{\kappa,\log}(x) \approx \frac{\mathcal{R}_{\text{clus}}^{\kappa}(x) - \mathcal{R}_{\text{clus}}^{\text{EM}}(x)}{\sigma_{\text{tot}}^{\kappa}(x)}$$

- stochastic data defining shape
- Gaussian with order-one width
- BNN and RE errors consistent
- per-cluster error conservative

# BNN and RE — uncertainties vs data spread



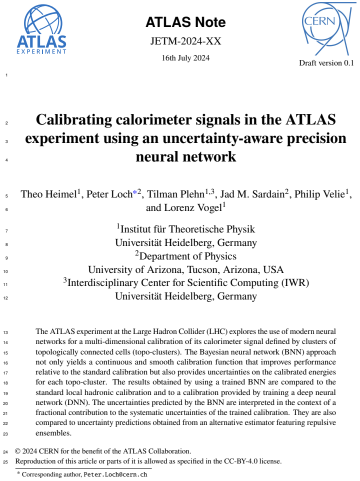

Central prediction plus uncertainty as pull (evaluated in $\log_{10} \mathcal{R}_{clus}$ space)

$$t_{tot}^{\kappa,\log}(x) \approx \frac{\mathcal{R}_{clus}^{\kappa}(x) - \mathcal{R}_{clus}^{EM}(x)}{\sigma_{tot}^{\kappa}(x)}$$

- stochastic data defining shape
- Gaussian with order-one width
- BNN and RE errors consistent
- per-cluster error conservative

# Outlook

# Outlook


**ATLAS Note**
JETM-2024-XX
16th July 2024
Draft version 0.1

**Calibrating calorimeter signals in the ATLAS experiment using an uncertainty-aware precision neural network**

Theo Heimel[1], Peter Loch*[2], Tilman Plehn[1,3], Jad M. Sardain[2], Philip Velie[1], and Lorenz Vogel[1]

[1]Institut für Theoretische Physik
Universität Heidelberg, Germany
[2]Department of Physics
University of Arizona, Tucson, Arizona, USA
[3]Interdisciplinary Center for Scientific Computing (IWR)
Universität Heidelberg, Germany

The ATLAS experiment at the Large Hadron Collider (LHC) 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 (DNN). 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 featuring 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.

* Corresponding author, Peter.Loch@cern.ch

## Modern BNNs for multi-dimensional topo-cluster calibration

- continuous and smooth calibration
- improved performance relative to LCW and DNN
- meaningful per-cluster systematics
- BNNs and REs: learned reliable uncertainties

## Next steps

- tune BNN performance
- boosted training to control remaining tails
- ATLAS paper in preparation... stay tuned!

*Thanks to Manuel Haußmann for introducing us to BNNs and REs!*

# What theorists can teach us about calorimeter calibration...

# References and further reading

## ML-based topo-cluster calibration


ATLAS Collaboration
*The application of neural networks for the calibration of topological cell clusters in the ATLAS calorimeters*
ATLAS PUB Note (2023)


## ML with uncertainties


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)


- S. Bollweg, M. Haußmann, G. Kasieczka, M. Luchmann, T. Plehn and J. Thompson
  *Deep-learning jets with uncertainties and more*
  SciPost Phys. 8, 006 (2020), arXiv:1904.10004 [hep-ph]
- 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]
- A. Butter, T. Heimel, S. Hummerich, T. Krebs, T. Plehn, A. Rousselot and S. Vent
  *Generative networks for precision enthusiasts*
  SciPost Phys. 14, 078 (2023), arXiv:2110.13632 [hep-ph]
- S. Badger, A. Butter, M. Luchmann, S. Pitz and T. Plehn
  *Loop amplitudes from precision networks*
  SciPost Phys. Core 6, 034 (2023), arXiv:2206.14831 [hep-ph]

# References and further reading


- A. Butter, B. M. Dillon, T. Plehn and L. Vogel
  *Performance versus resilience in modern quark-gluon tagging*
  SciPost Phys. Core 6, 085 (2023), arXiv:2212.10493 [hep-ph]


## Repulsive ensembles (REs)


- 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:** The dataset consists of topo-clusters reconstructed in MC simulations of full proton-proton collision events at $\sqrt{s} = 13\,\text{TeV}$ (LHC Run 2) with multi-jet final states

| **category** | **symbol** | **description / comment** |
|---|---|---|
| kinematics | $E^{EM}_{clus}$, $y^{EM}_{clus}$ | cluster signal and rapidity at the EM energy scale |
| signal strength | $\zeta^{EM}_{clus}$ | signal significance |
| timing<br>time structure | $t_{clus}$<br>$\mathrm{Var}_{clus}(t_{cell})$ | signal timing<br>variance of the cell-time distribution in the cluster |
| shower depth | $\lambda_{clus}$<br>$\lvert\vec{c}_{clus}\rvert$ | distance of the CoG from the calorimeter front face<br>distance of the CoG from the nominal vertex |
| shower shape | $f_{emc}$<br>$\langle\rho_{cell}\rangle$, $p_T D$<br>$\langle\mathfrak{m}^2_{long}\rangle$, $\langle\mathfrak{m}^2_{lat}\rangle$ | energy fraction in the EM calorimeter (EMC)<br>cluster signal density and signal compactness<br>energy dispersion along/perpendicular to main cluster axis |
| topology | $f_{iso}$ | cluster isolation measure |
| pile-up | $N_{PV}$<br>$\mu$ | number of reconstructed primary vertices<br>number of pile-up interactions per bunch crossing |

# Dataset — feature pre-processing

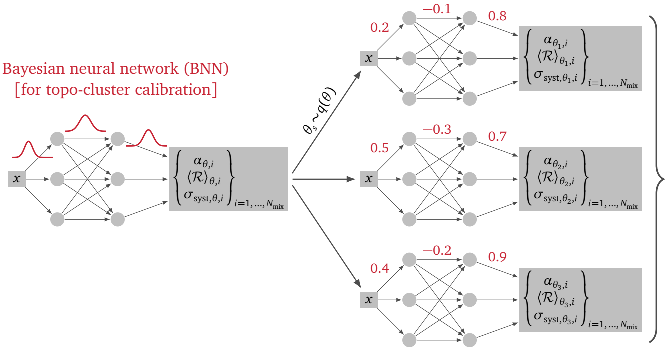
**Table 2:** To limit the numerical value range of the network inputs and to create more appropriate (peaked) feature distributions, the features are pre-processed

| **transformation** | **features** |
|---|---|
| $\log_{10}$ standardization | $E^{\mathrm{EM}}_{\mathrm{clus}}$, $\zeta^{\mathrm{EM}}_{\mathrm{clus}}$, $\mathrm{Var}_{\mathrm{clus}}(t_{\mathrm{cell}})$, $\lambda_{\mathrm{clus}}$, $\langle\rho_{\mathrm{cell}}\rangle$ |
| linear standardization | $y^{\mathrm{EM}}_{\mathrm{clus}}$, $\lvert\vec{c}_{\mathrm{clus}}\rvert$, $f_{\mathrm{emc}}$, $\langle\mathfrak{m}^2_{\mathrm{long}}\rangle$, $\langle\mathfrak{m}^2_{\mathrm{lat}}\rangle$, $p_{\mathrm{T}}D$, $f_{\mathrm{iso}}$, $N_{\mathrm{PV}}$, $\mu$ |
| maximum-absolute normalization | $t_{\mathrm{clus}}$ |
| $\log_{10}$ transformation | $\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}$ |

# BNN — network architecture

Bayesian neural network (BNN)
[for topo-cluster calibration]

ensemble of $S$ networks by sampling $\theta_s$
from learned variational distributions $q(\theta)$




evaluation: maximum-likelihood value
and systematic and statistical uncertainties
[for Gaussian mixture likelihood]

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

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

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

with:

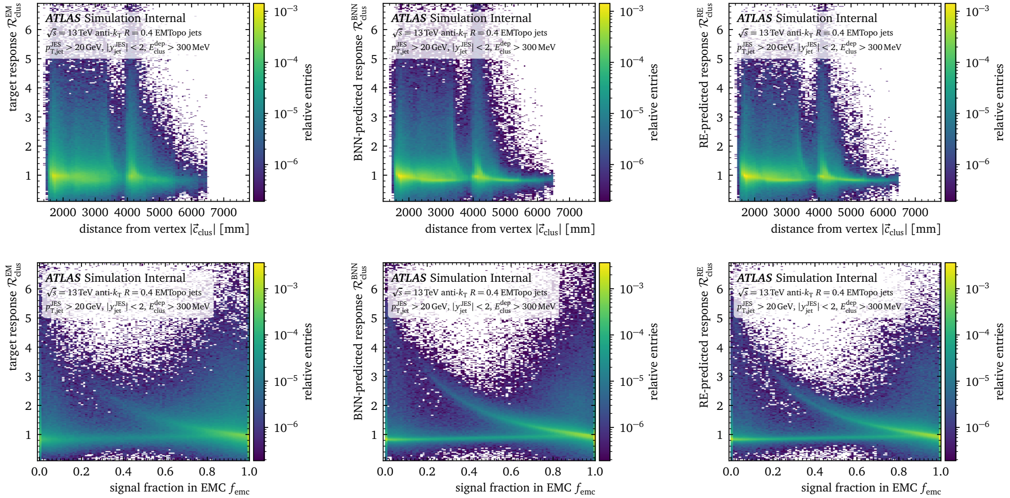
$$\langle \mathcal{R} \rangle_{\theta_s} = \sum_{i=1}^{N_{\text{mix}}} \alpha_{\theta_s,i} \langle \mathcal{R} \rangle_{\theta_s,i} \quad \text{and} \quad \langle \mathcal{R} \rangle = \frac{1}{S} \sum_{s=1}^{S} \langle \mathcal{R} \rangle_{\theta_s}$$

# Dataset — feature pre-processing

**Table 3:** BNN setup for the three-mode Gaussian mixture likelihood

| **hyper-parameter** | **BNN architecture and setup** |
|---|---|
| likelihood loss | Gaussian mixture model (GMM) |
| number of modes (mixture components $N_{\text{mix}}$) | 3 (i.e. 9 output nodes) |
| number of layers and nodes per layer | 4 hidden layers with $\{64, 64, 64, 64\}$ |
| activation functions | ReLU (inner layers) and none (last layer) |
| prediction | maximum-likelihood value |
| optimizer and learning rate (LR) | ADAM with $\text{LR} = 10^{-4}$ |
| learning-rate scheduler | STEPLR, epochs $\{25, 100\}$, $\gamma = 0.1$ |
| number of training epochs | 150 |
| batch size for training (testing) | 4096 (512) |
| dataset sizes for training, validation, testing | $\{8.7\text{M}, 500\text{k}, 5.3\text{M}\}$ |
| re-sampling for testing (Monte-Carlo samples $S$) | 50 times |

# EM, BNN and RE — response vs features



# DNN, BNN and RE — signal linearity vs features