The Application of Neural Networks for the Calibration of ATLAS Calorimeter Signals

Peter Loch (o.b.o. the ATLAS Collaboration)
Department of Physics, University of Arizona
Tucson, Arizona 85721, USA

THE UNIVERSITY OF ARIZONA
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

ML4Jets 2023
Hamburg - 6-10 November

ATLAS EXPERIMENT

This talk

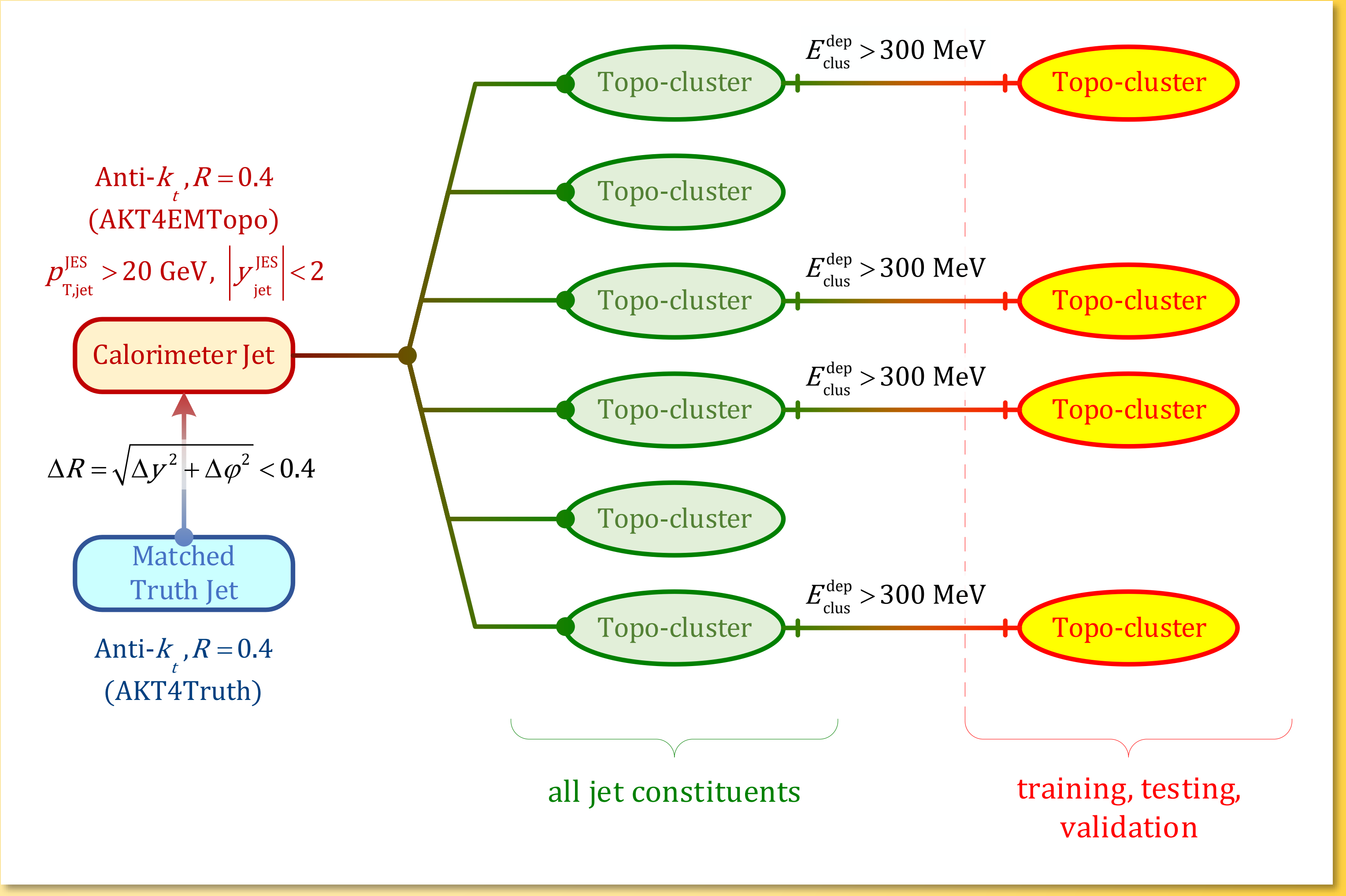
- Motivation
  - Local calibration
  - This project
- Signal feature selection and network designs
  - Sample collection
  - Signal features useful for calibration
  - DNN & BNN networks
- Performance evaluations
  - Prediction power
  - Signa linearity and (local) resolution
- Conclusion
  - Future plans
- Reference
  - All plots (and many more) can be found in ATLAS public note ATL-PHYS-PUB-2023-019

ML-based Local Calibration

- Local calorimeter response and resolution
  - Improved constituent inter-calibration in combined tracking/calorimeter final state reconstruction
    - Better calibrated neutral response in jets – improvement in scale and resolution for all jets, sub-structure variable measurements, measurement of full hadronic recoil (non-jet context), measurement of hadronic event shapes …
  - Replacement for present-day LCW with ML-based approach as a local calibration for topo-clusters
    - LCW: multi-dimensional binned look-up tables → ML: smooth multidimensional calibration functions, no steep steps at bin edges, …
    - LCW: loss of correlations due to average scale factors in bins → ML: exploitation and preservation of correlations, better resolution
- Intentional limitations of approach
  - Looking for practical application to be applied to collision data
    - Extract topo-clusters for training and testing from calorimeter jets in fully simulated events with Run 2 level pile-up
    - Use of cluster moments (constructed features) allows recalibrating at derivation level – all needed data is in the AOD
    - Quick adaptation to changing collision environments (e.g., pile-up) and reconstruction cuts – newly trained networks can be applied at derivation level (data preparation/extraction for physics)
    - No need to go back to full (Tier0) reconstruction, no detailed information (calorimeter cells) needed
  - Single-step approach
    - No dedicated (learned) classification prior to regression on topo-cluster response

Signal Source from MC Simulations

- Full simulation of detector signals in pp collision jet production final states with LHC Run 2 pile-up
  - Jets in central detector region only
  - Need to match generated truth particle jet
- Topo-cluster extracted from jets
  - $E_{\text{clus}}^{\text{dep}} > 300$ MeV
  - Randomly selected for independent training, validation, and test samples
  - No jet-specific information/features used (jet context is fully removed)



Local Hadronic Calibration in ATLAS

Local topo-cluster calibration with Monte Carlo simulations

Standard (LCW) local hadronic calibration

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

EM-likelihood | hadronic calibration | outside energy losses

$$E_{\text{clus}}^{\text{had}} = \sum_{\text{cells } i \text{ in cluster}} w_{\text{cell},i}^{\text{geo}} \cdot \left(\mathcal{P}_{\text{clus}}^{\text{EM}} + (1 - \mathcal{P}_{\text{clus}}^{\text{EM}}) \cdot w_{\text{cell},i}^{\text{cal}}\right) \times E_{\text{cell},i} \xrightarrow{\text{measures}} \underbrace{E_{\text{clus}}^{\text{dep}}}_{}$$

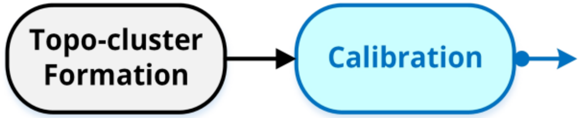
energy deposited in cluster (from MC)

$\mathcal{P}_{\text{clus}}^{\text{EM}}$ & $w_{\text{cell}}^{\text{cal}}$ from *look-up tables* employing cluster moments & cell signal features

1. **Classification** – look up likelihood for cluster being of EM nature $0 \leq \mathcal{P}_{\text{clus}}^{\text{EM}} \leq 1$
2. **Hadronic calibration** – apply cell signal weights to correct hadronic signals to EM scale ($e/h > 1 \xrightarrow{\text{calibrate to}} e/h = 1$)
3. **Out-of-cluster corrections** – correct for energy lost due to cells missed in cluster formation
4. **Dead material corrections** – correct for energy losses in inactive material in the proximity of the topo-cluster

Machine-learning-based (ML-based) local calibration

(proof of concept study)



learned calibration

$$E_{\text{clus}}^{\text{ML}} = C_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right) \cdot E_{\text{clus}}^{\text{EM}} \xrightarrow{\text{measures}} E_{\text{clus}}^{\text{dep}}$$

energy deposited in cluster (from MC)

$E_{\text{clus}}^{\text{ML}} \mathrel{\hat{=}} E_{\text{clus}}^{\text{had}}$ — $E_{\text{clus}}^{\text{EM}}$: *basic cluster signal (EM scale)*

$$\mathcal{R}_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right) \xrightarrow{\text{predicts}} \mathcal{R}_{\text{clus}}^{\text{EM}} = \frac{E_{\text{clus}}^{\text{EM}}}{E_{\text{clus}}^{\text{dep}}} \xrightarrow{\text{yields}} C_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right) = \frac{1}{\mathcal{R}_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right)}$$

1. **No explicit classification** – variables used to determine $\mathcal{P}_{\text{clus}}^{\text{EM}}$ added to feature set $\mathfrak{O}_{\text{clus}}^{\text{ML}}$
2. **Hadronic calibration** – apply scale factor to cluster signal $E_{\text{clus}}^{\text{EM}}$ to measure $E_{\text{clus}}^{\text{dep}}$

Calibration from response predictions $\mathcal{R}_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right)$ obtained by regression fits $\mathcal{R}_{\text{clus}}^{\text{ML}}\left(\mathfrak{O}_{\text{clus}}^{\text{ML}}\right) \mapsto \mathcal{R}_{\text{clus}}^{\text{EM}}$ employing feature set $\mathfrak{O}_{\text{clus}}^{\text{ML}}$

Feature Set Composition

❖ Focus on observables contributing to response

A. Deposited energy represented by signal (in a complex way)

B. Detector geometry – signal characteristics of calorimeter sub-systems: variations in level of signal non-compensation ($e/h > 1$), absorption power/leakage, energy sharing around inactive regions …

C. Shower development – differences between EM and HAD showers: starting point, size, spread and compactness …

D. Intrinsic shower fluctuations – variations in the shower development of (hadronic) showers

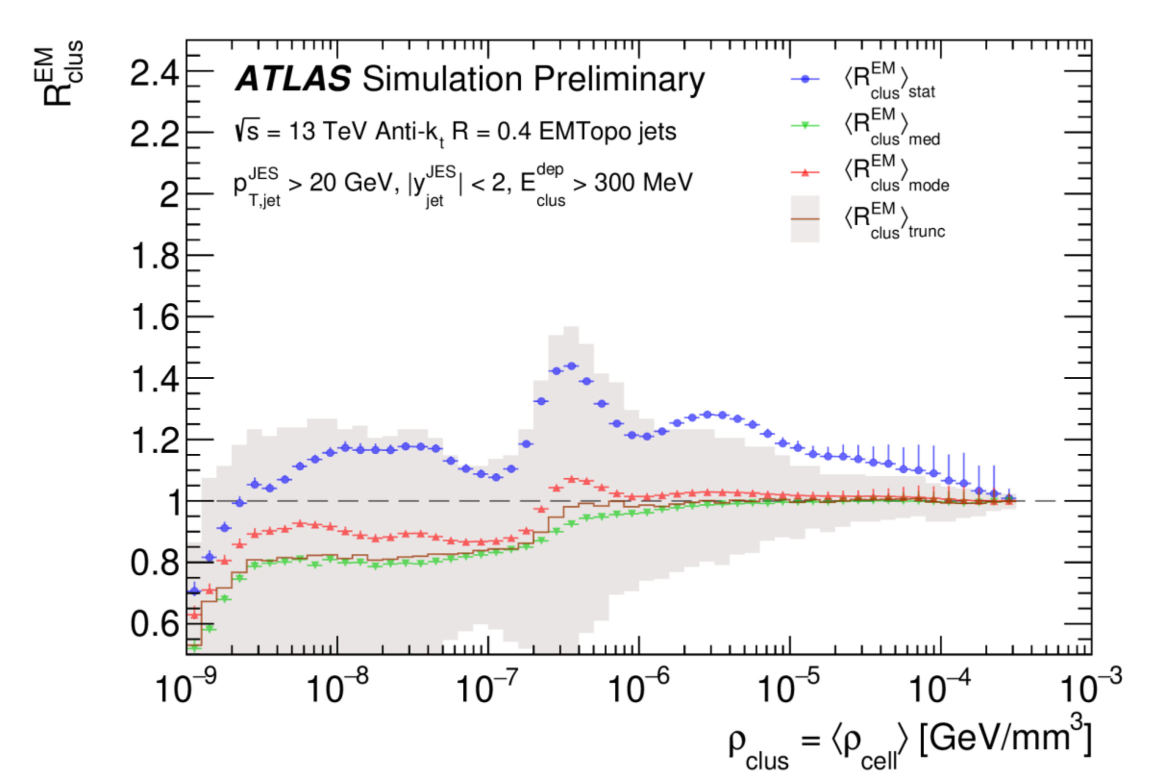
E. Signal strength and relevance – signal significance measured by signal-over-noise

F. Collision environment – effects of event topology/nearby signals (isolation) and pile-up on the topo-cluster signal

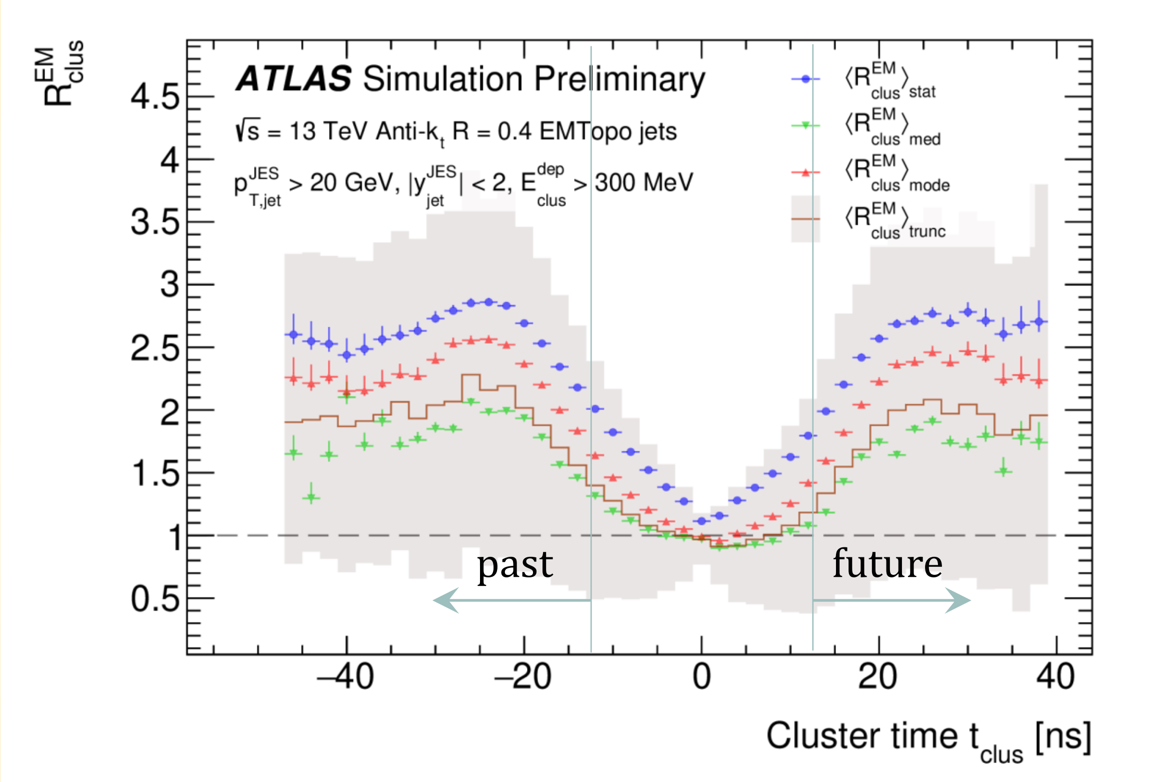
Category	Symbol	LCW	Comment
kinematics	$E^{\mathrm{EM}}_{\mathrm{clus}}$	yes	signal at the electromagnetic energy scale (A)
	$y^{\mathrm{EM}}_{\mathrm{clus}}$	yes	rapidity at the electromagnetic energy scale (B)
signal strength timing	$\zeta^{\mathrm{EM}}_{\mathrm{clus}}$	no	signal significance (E)
	t_{clus}	no	signal timing (C,D,F)
	$\mathrm{Var}_{\mathrm{clus}}(t_{\mathrm{cell}})$	no	variance of t_{cell} distribution (D,F)
shower depth shower shape compactness	λ_{clus}	yes	distance of centre-of-gravity from calorimeter front face (C,D)
	$\lvert\vec{c}_{\mathrm{clus}}\rvert$	no	distance of centre-of-gravity from nominal vertex (C,D)
	f_{emc}	no	fraction of energy in electromagnetic calorimeter (C)
	$\langle\rho_{\mathrm{cell}}\rangle$	yes	cluster signal density measure (C,D)
	$\langle\mathfrak{m}^2_{\mathrm{long}}\rangle$	no	energy dispersion along main cluster axis (C)
	$\langle\mathfrak{m}^2_{\mathrm{lat}}\rangle$	no	energy dispersion perpendicular to main cluster axis (C)
	$p_{\mathrm{T}}D$	no	signal compactness measure (C,D)
topology	f_{iso}	no	cluster isolation measure (F)
pile-up	N_{PV}	no	number of reconstructed primary vertices (F)
	μ	no	number of interactions per bunch crossing (F)

$$\mathfrak{O}^{\mathrm{ml}}_{\mathrm{clus}} = \{\underbrace{E^{\mathrm{EM}}_{\mathrm{clus}}, y^{\mathrm{EM}}_{\mathrm{clus}}}_{\text{kinematics}}, \overbrace{\zeta^{\mathrm{EM}}_{\mathrm{clus}}, t_{\mathrm{clus}}, \mathrm{Var}_{\mathrm{clus}}(t_{\mathrm{cell}})}^{\text{signal strength and timing}}, \underbrace{\lambda_{\mathrm{clus}}, \lvert\vec{c}_{\mathrm{clus}}\rvert, \langle\rho_{\mathrm{cell}}\rangle, \langle\mathfrak{m}^2_{\mathrm{long}}\rangle, \langle\mathfrak{m}^2_{\mathrm{lat}}\rangle, p_{\mathrm{T}}D, f_{\mathrm{emc}}}_{\text{shower location (depth), shapes and compactness}}, \overbrace{f_{\mathrm{iso}}}^{\text{topology (isolation)}}, \underbrace{N_{\mathrm{PV}}, \mu}_{\text{event/pile-up}}\}$$

Response Dependence on Features



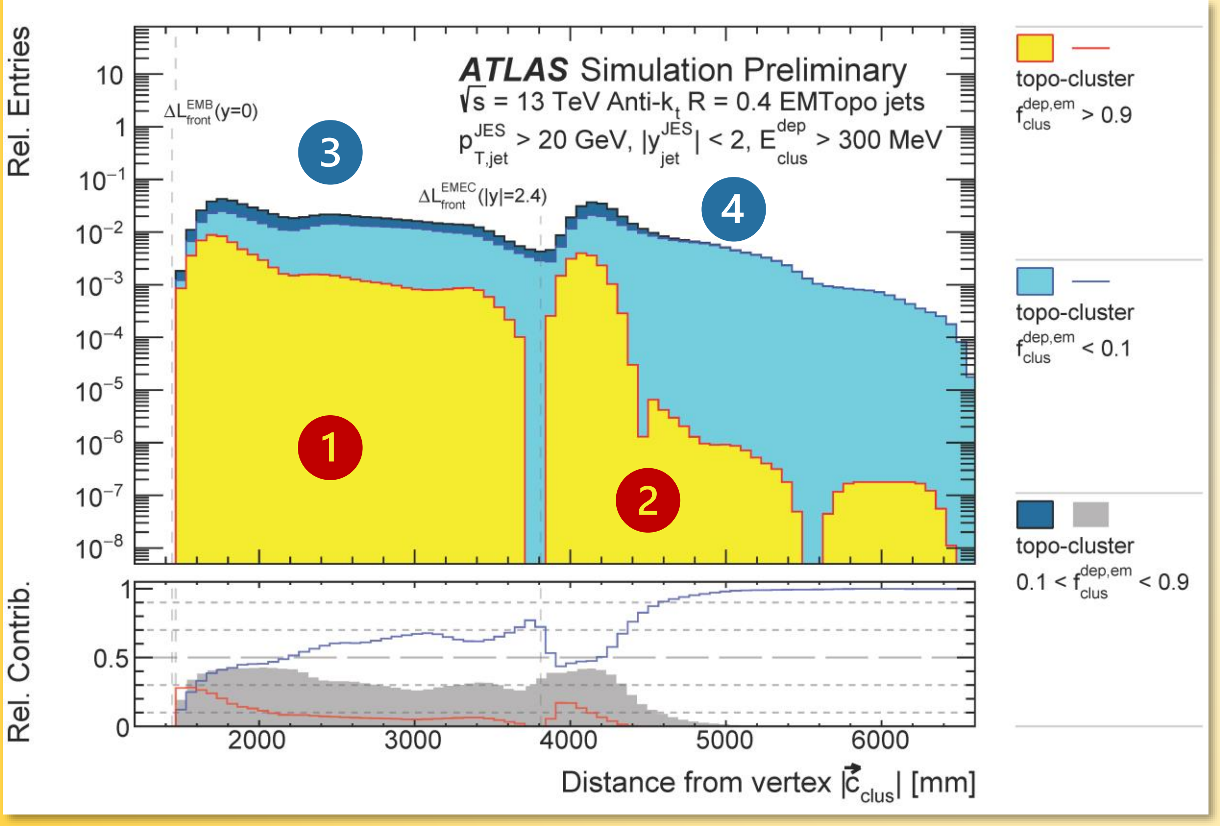
Measure for tendency of centrality influences ML setup → loss function definition

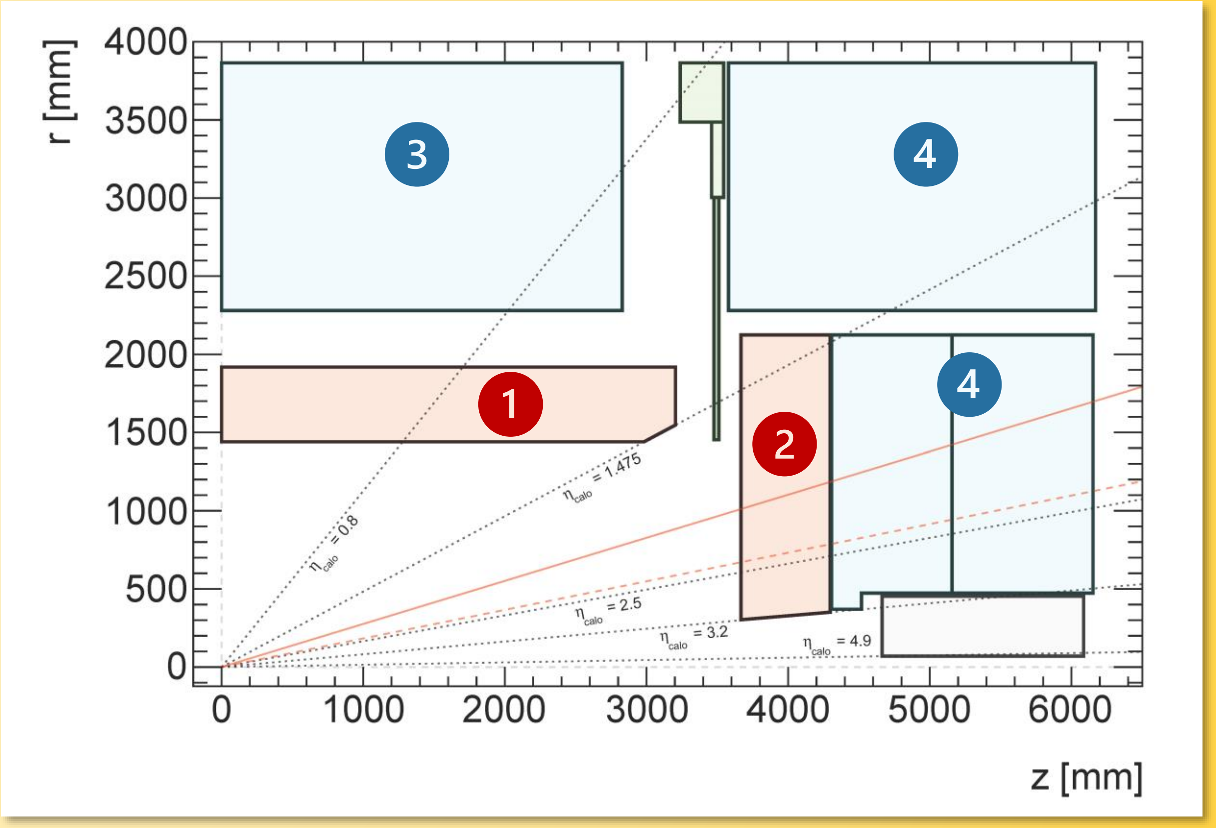


Cluster timing is affected by (out-of-time) pile-up → response increased by additional signals from nearby past bunch crossing and the following one

Challenge: Feature Distributions

- Example: complex distribution of topo-cluster distance from nominal vertex



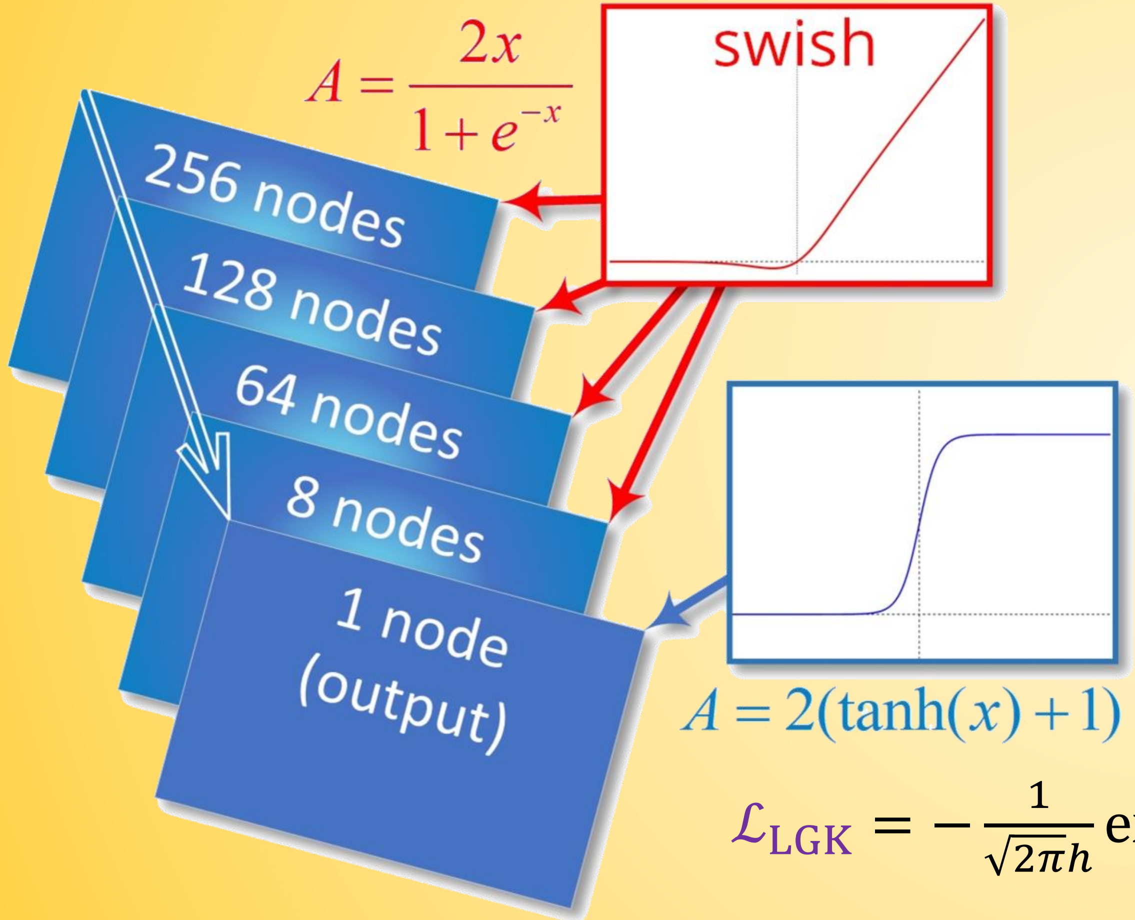


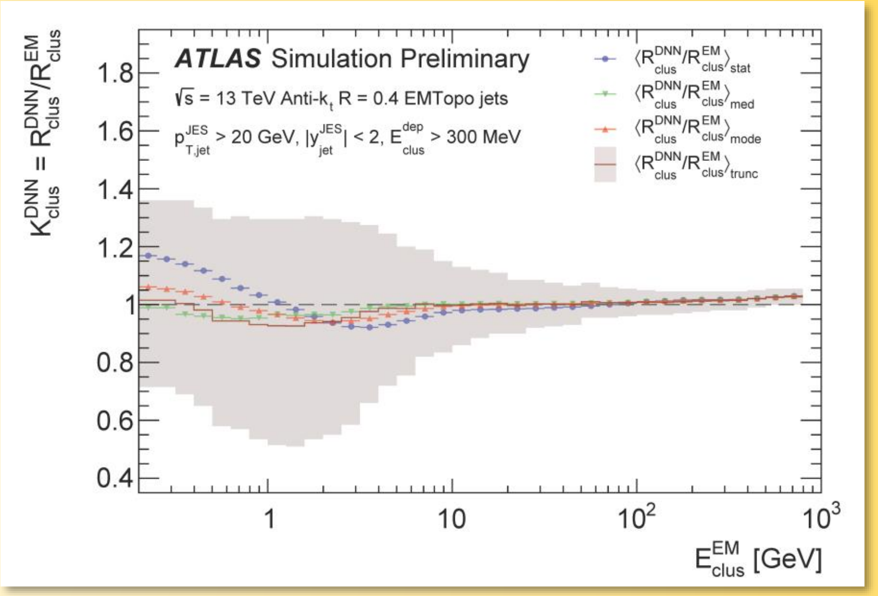
predominantly $e^{\pm}/\gamma$ | mixed | predominantly hadronic

contribution to deposited energy

Neural Network Designs: DNN

- ❖ Highly tuned configuration





Loss function trains the mode:

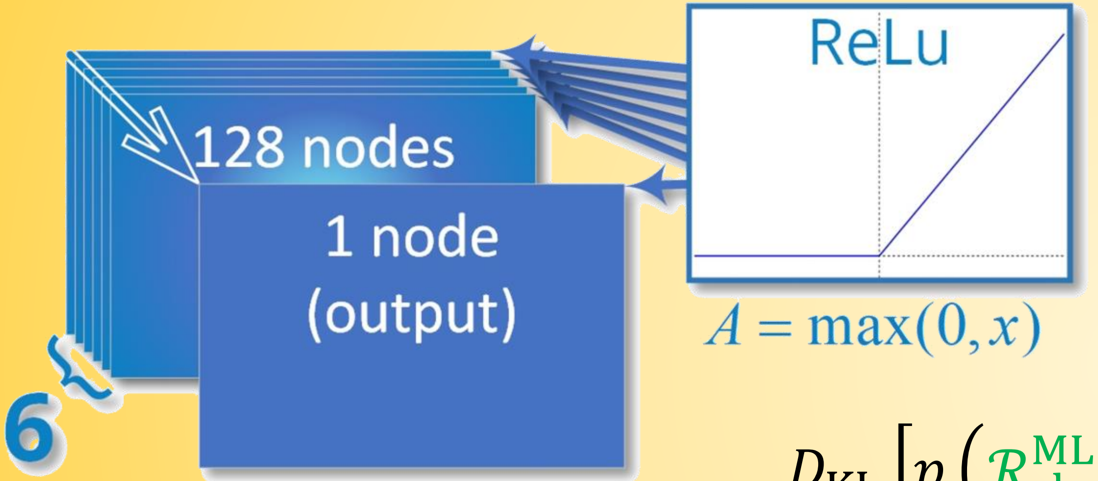
1st pass: $h = 0.1, \alpha = 0.05$

2nd pass (seeded with 1st pass model)**:** $\alpha = 0$

$$\mathcal{L}_{\mathrm{LGK}} = -\frac{1}{\sqrt{2\pi h}} \exp\left[\frac{1}{2h}\left(\frac{\mathcal{R}^{\mathrm{ML}}_{\mathrm{clus}}(\mathfrak{O}^{\mathrm{ML}}_{\mathrm{clus}})}{\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}} - 1\right)^2\right] + \alpha \left|\frac{\mathcal{R}^{\mathrm{ML}}_{\mathrm{clus}}(\mathfrak{O}^{\mathrm{ML}}_{\mathrm{clus}})}{\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}} - 1\right|$$

Neural Network Designs: BNN

❖ First attempt*



Negative log-likelihood loss function with regularization by (reverse) Kullback-Leibler (KL) divergence D_{KL} *models* $\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}$ *distribution*

$$D_{\mathrm{KL}}\left[p\left(\mathcal{R}^{\mathrm{ML}}_{\mathrm{clus}}\left(\mathfrak{D}^{\mathrm{ML}}_{\mathrm{clus}}\right)\right), q\left(\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}\right)\right] = \left\langle \log \frac{p\left(\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}\right)}{q\left(\mathcal{R}^{\mathrm{ML}}_{\mathrm{clus}}\left(\mathfrak{D}^{\mathrm{ML}}_{\mathrm{clus}}\right)\right)} \right\rangle_{q\left(\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}\right)}$$

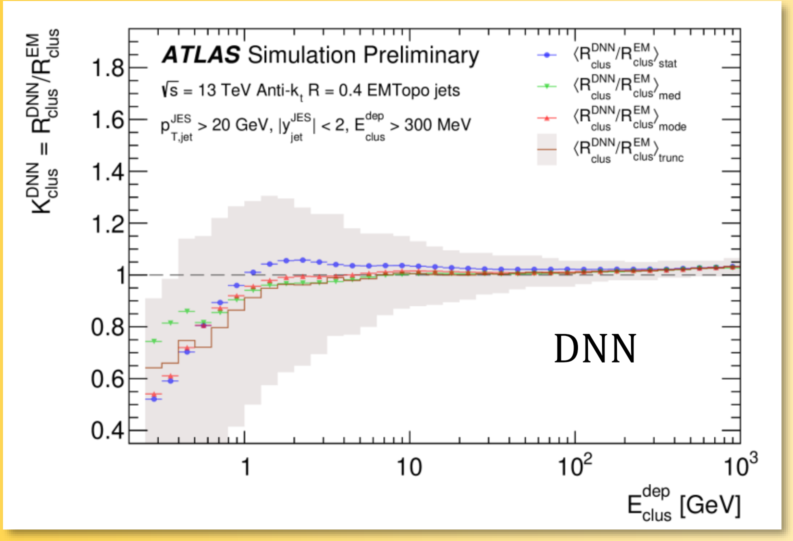
$$\mathcal{L}_{\mathrm{BNN}} \sim -\log\left(p_{\mathrm{train}} = \sum_{i=1}^{6} \mathcal{G}_i\left(\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}, \mathfrak{D}^{\mathrm{ML}}_{\mathrm{clus}}; \mu_i, \sigma_i\right)\right) + D_{\mathrm{KL}}\left[p\left(\mathcal{R}^{\mathrm{ML}}_{\mathrm{clus}}\left(\mathfrak{D}^{\mathrm{ML}}_{\mathrm{clus}}\right)\right), q\left(\mathcal{R}^{\mathrm{EM}}_{\mathrm{clus}}\right)\right]$$

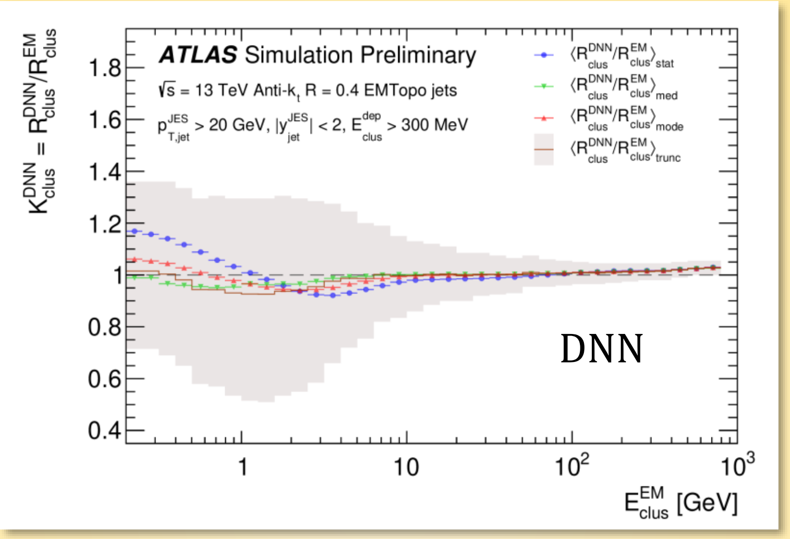
↪ uncertainties (1) for model and (2) from training statistics → under further investigation

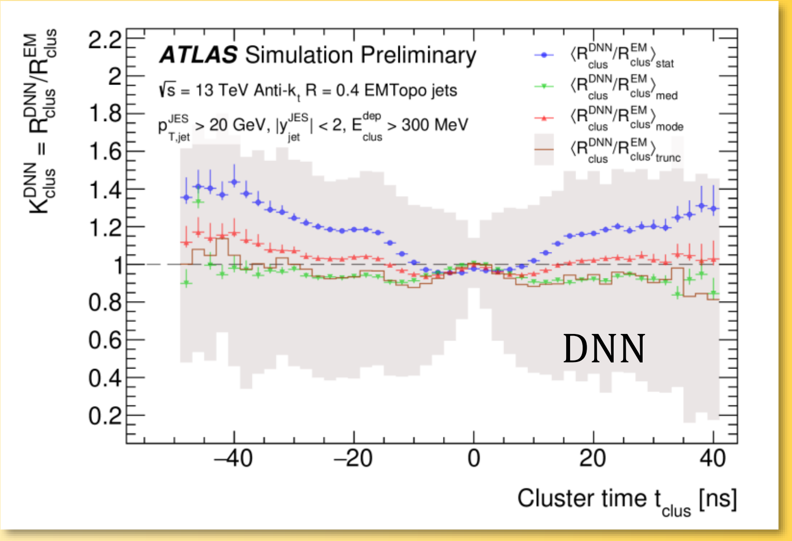
*with many thanks to Tilman Plehn & Michel Luchmann for providing the code and lots of advice

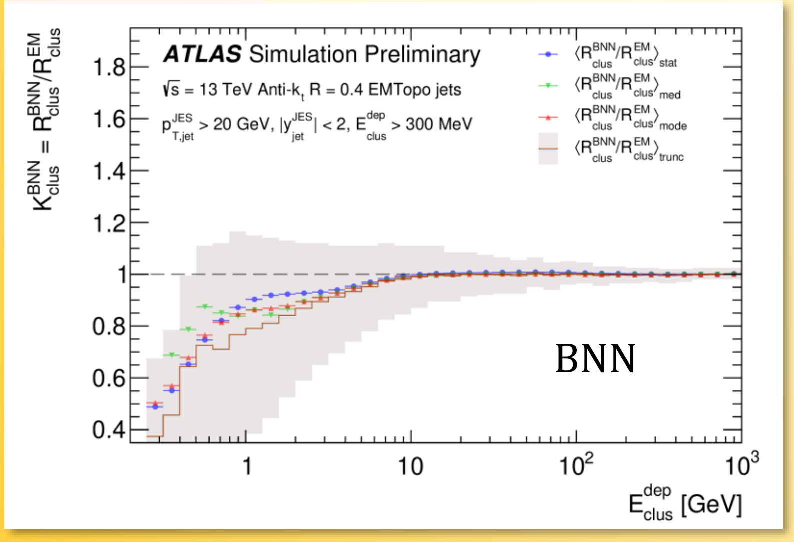
Prediction Power

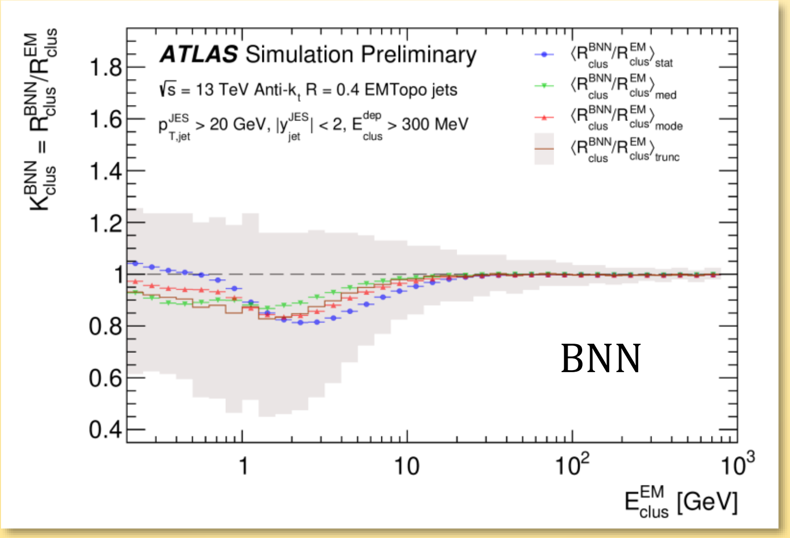
$$K_{\mathrm{clus}}^{\mathrm{DNN(BNN)}} = R_{\mathrm{clus}}^{\mathrm{DNN(BNN)}} / R_{\mathrm{clus}}^{\mathrm{EM}}$$

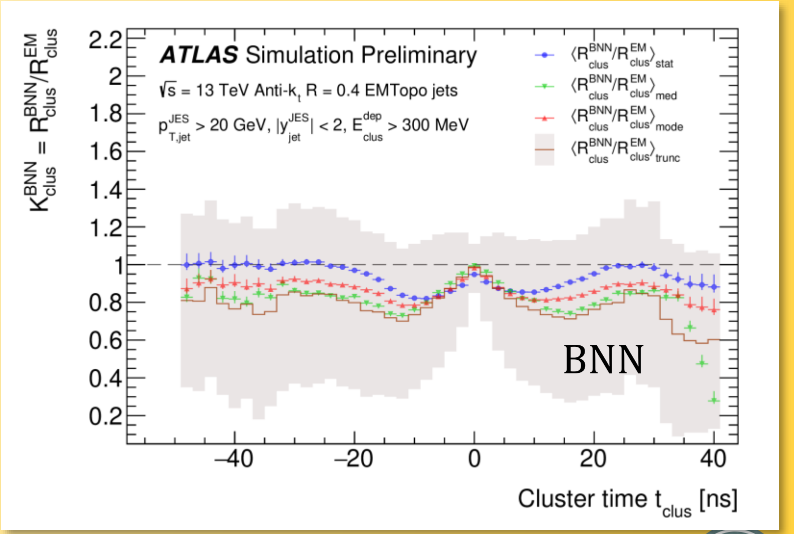






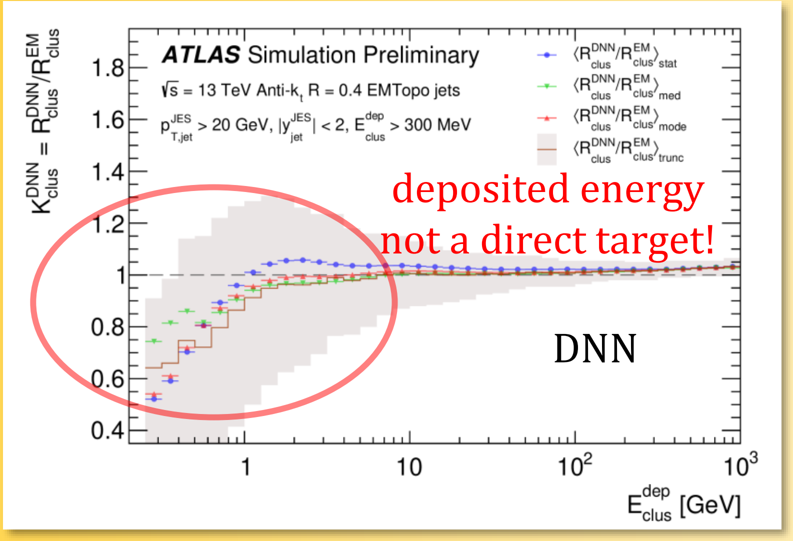


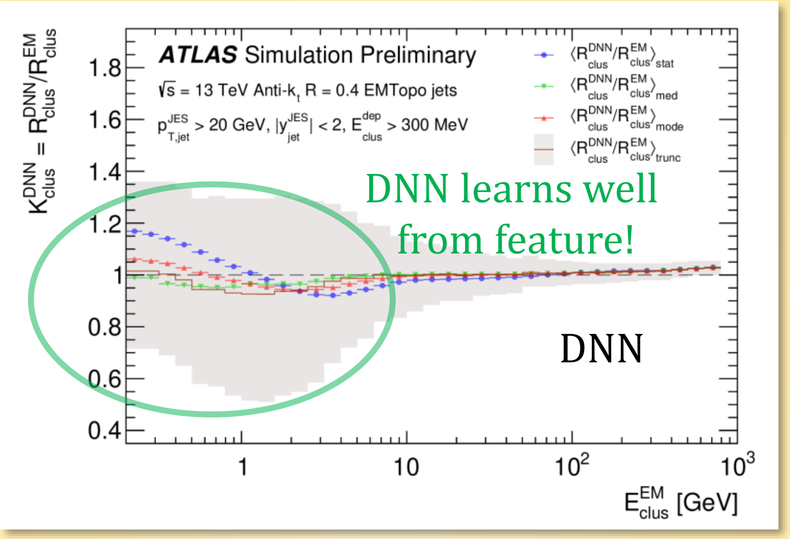


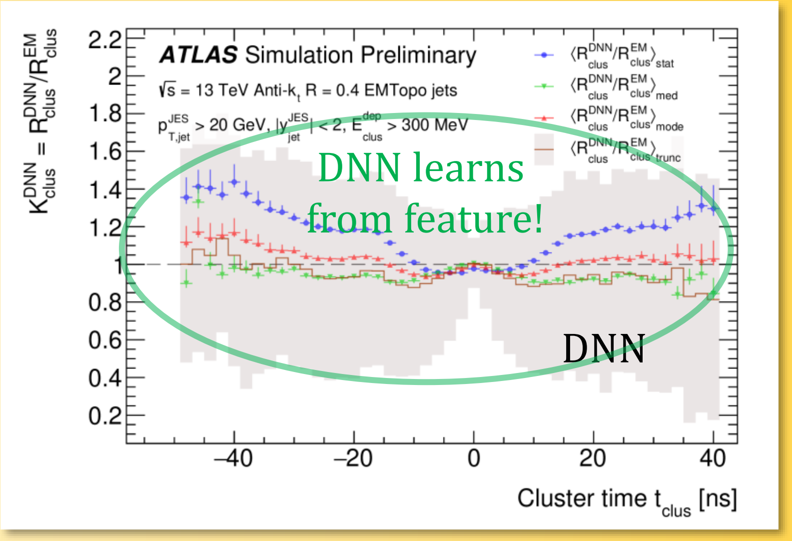


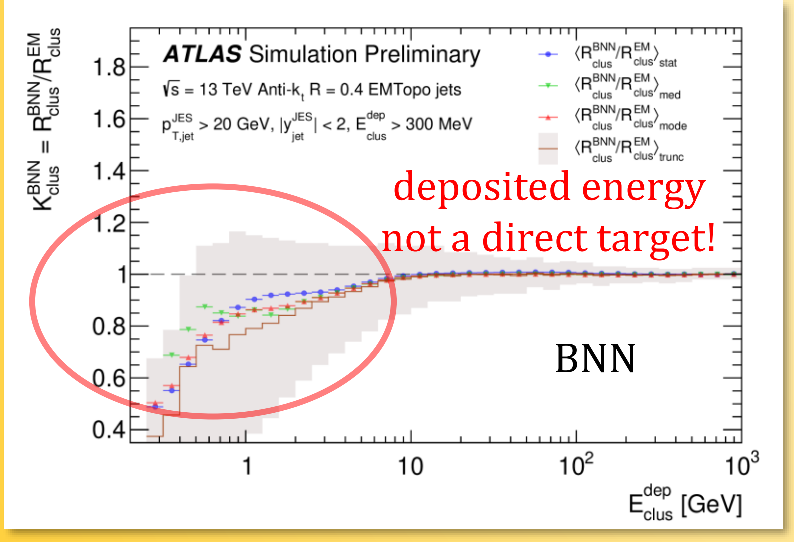
Prediction Power

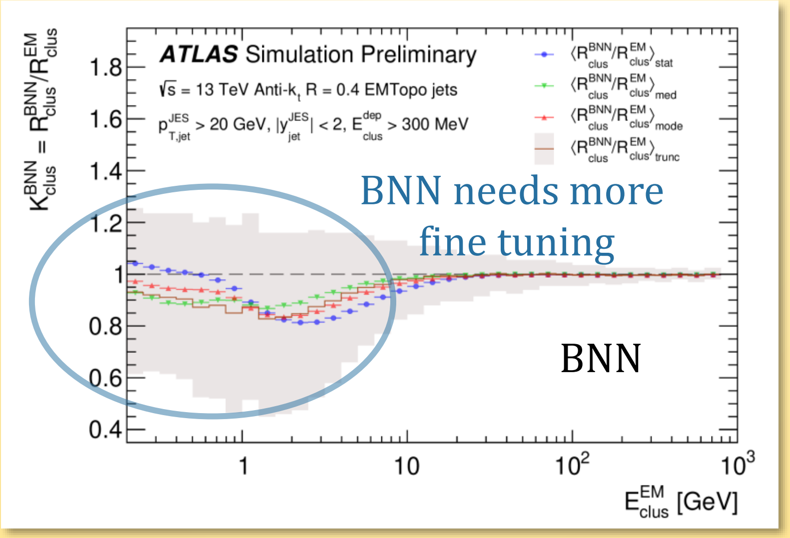
$$K_{\text{clus}}^{\text{DNN(BNN)}} = R_{\text{clus}}^{\text{DNN(BNN)}} / R_{\text{clus}}^{\text{EM}}$$

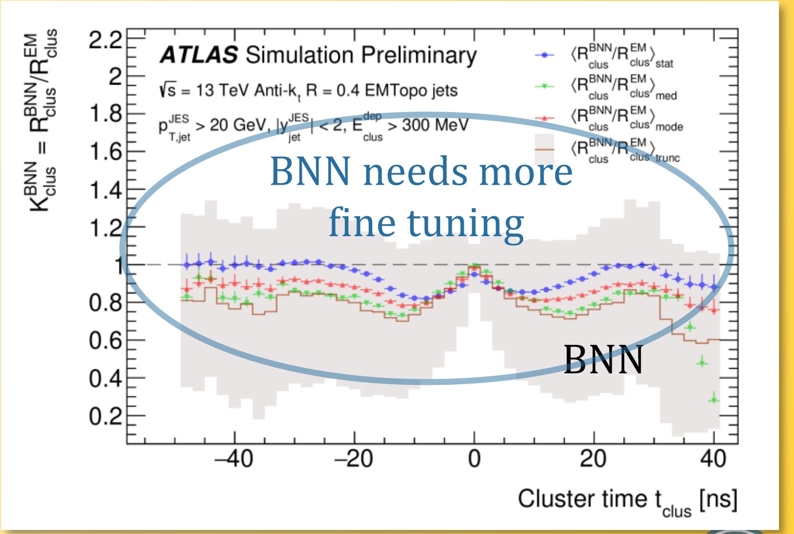




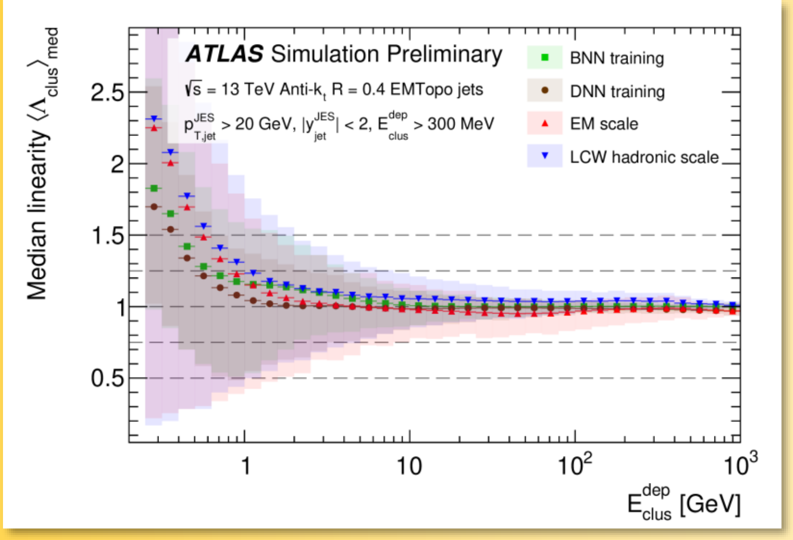


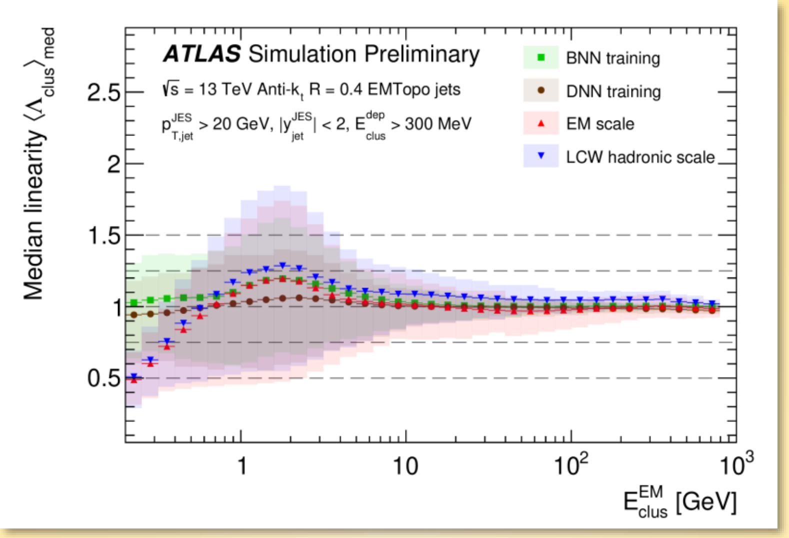


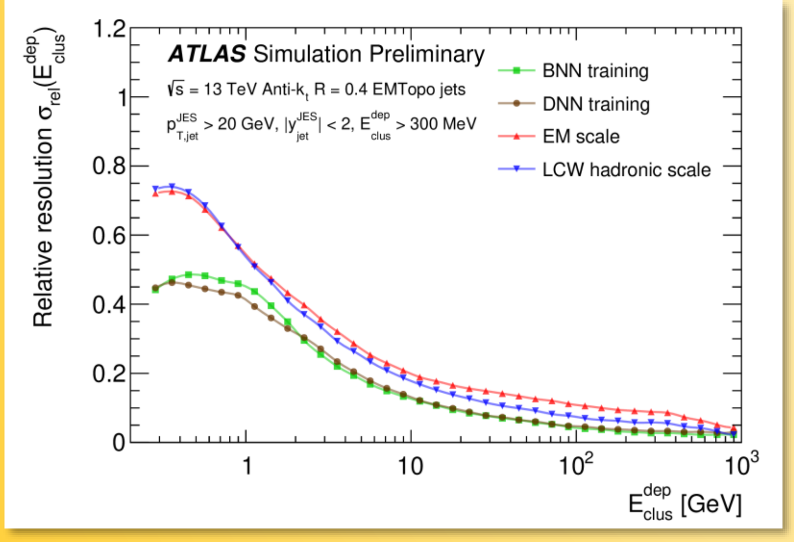


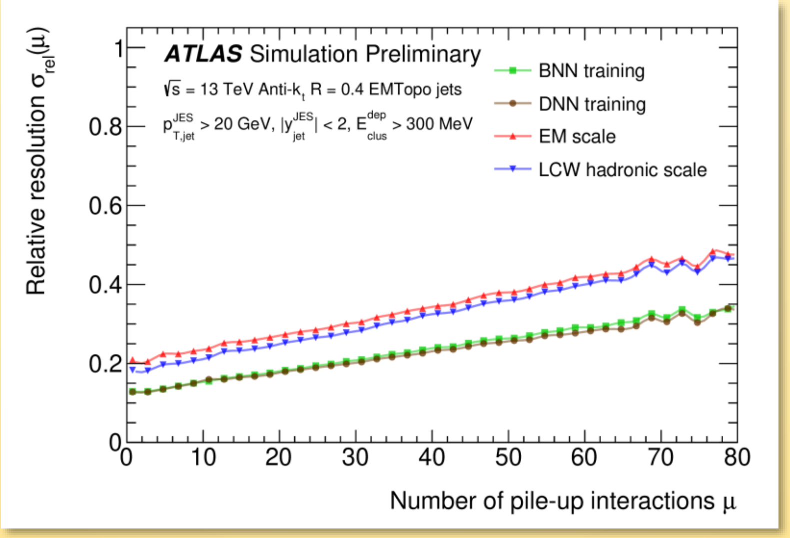


Signal Linearity & Resolution







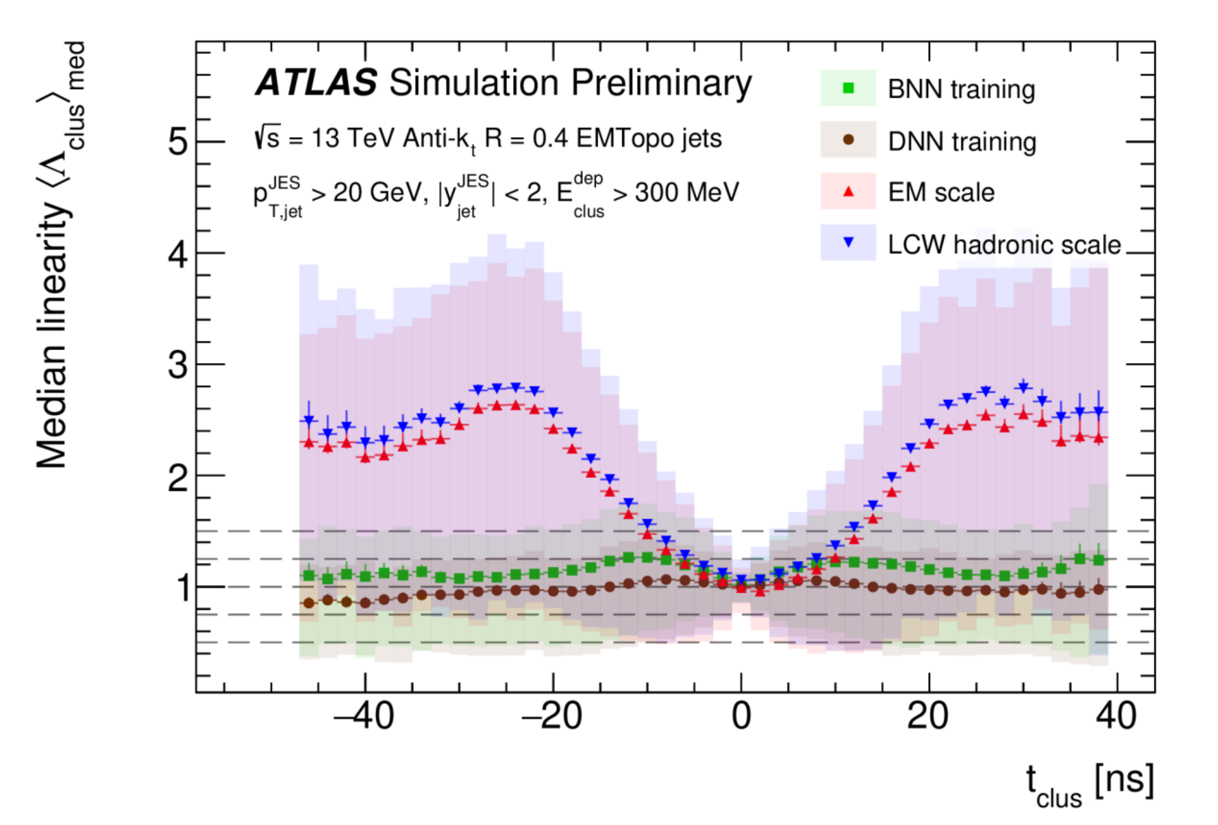


- **Linearity**
  - $\langle \Lambda_{\mathrm{clus}} \rangle_{\mathrm{med}} = \left\langle E_{\mathrm{clus}}^{\mathrm{reco}} / E_{\mathrm{clus}}^{\mathrm{dep}} \right\rangle_{\mathrm{med}}$
  - $E_{\mathrm{clus}}^{\mathrm{reco}} \in \{E_{\mathrm{clus}}^{\mathrm{EM}}, E_{\mathrm{clus}}^{\mathrm{LCW}}, E_{\mathrm{clus}}^{\mathrm{DNN}}, E_{\mathrm{clus}}^{\mathrm{BNN}}\}$
- **Resolution**
  - $\sigma_{\mathrm{rel}} = \mathrm{IQR}_{\mathrm{clus}}^{\Lambda_{\mathrm{clus}}} / (2 \cdot \langle \Lambda_{\mathrm{clus}} \rangle_{\mathrm{med}})$
- **Findings**
  - Improved linearity as a function of features for DNN and BNN
  - Present day LCW calibration agnostic to some features (not $E_{\mathrm{clus}}^{\mathrm{EM}}$!)
  - Local energy resolution significantly improved
  - Pile-up effects on resolution generally reduced – slight improvement in slope as well

Maybe there is more?

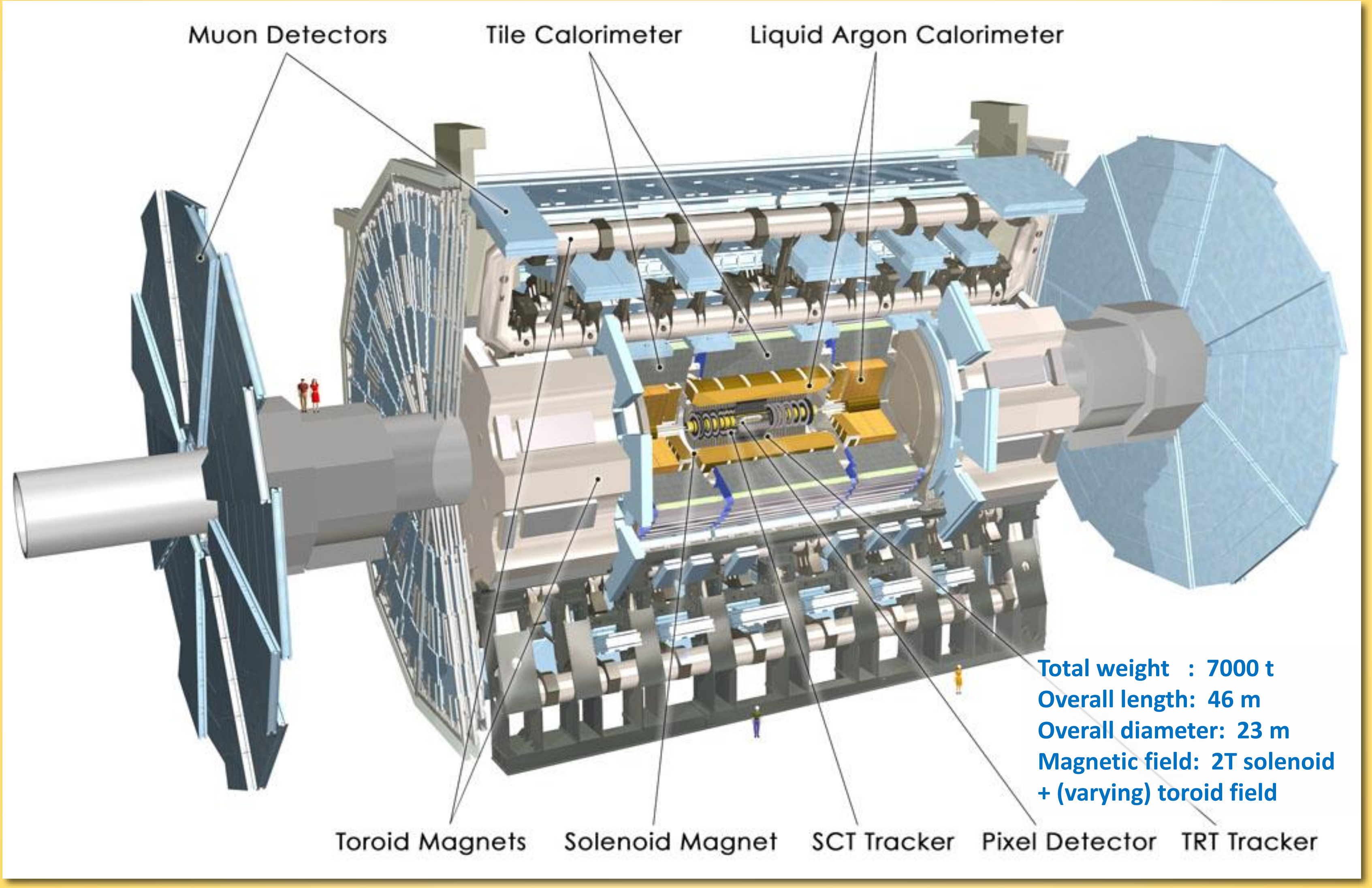
- Out-of-time pile-up mitigation
  - DNN/BNN learn time dependence of response very well – significantly improved signal linearity
  - Yields “built-in” correction for pile-up
  - Can provide basis for classification as well (?)
- Other expectations for performance improvements
  - Tests in full jet context still outstanding – what happens to pure ($E_{\text{clus}}^{\text{dep}} = 0$) and pile-up dominated topo-clusters ($E_{\text{clus}}^{\text{dep}} < 300$ MeV)
  - Local resolution improvement promising for (softer) hadronic recoil reconstructions
- Exploration of BNN
  - Understanding of uncertainty predictions – contribution to “bottoms-up” systematics
  - Now an ATLAS project with help from theorists

Very promising first results for (specific) topo-clusters found in realistic *pp* collision environment!



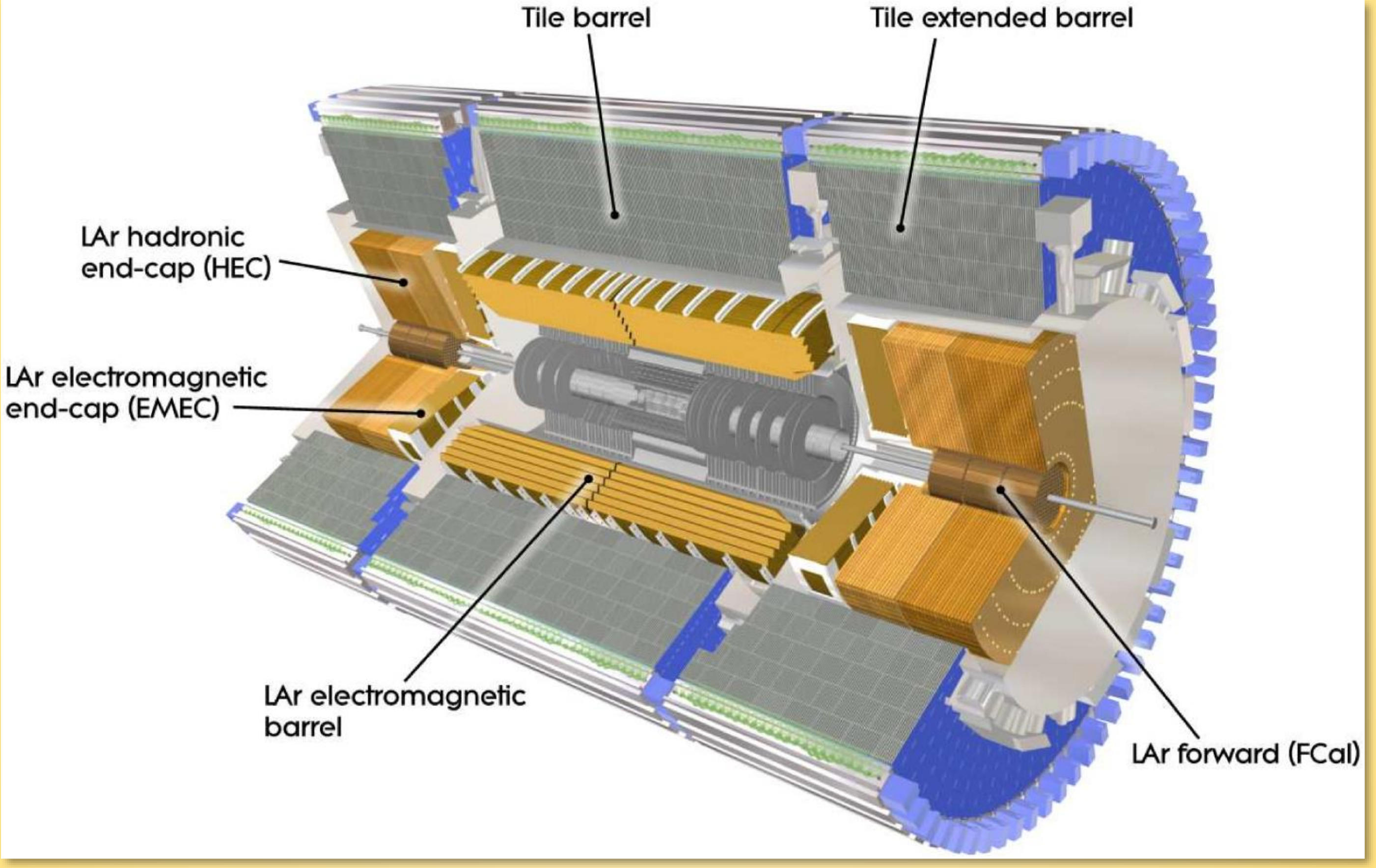
Backup & Extra Slides

Moving on to the Experiment…

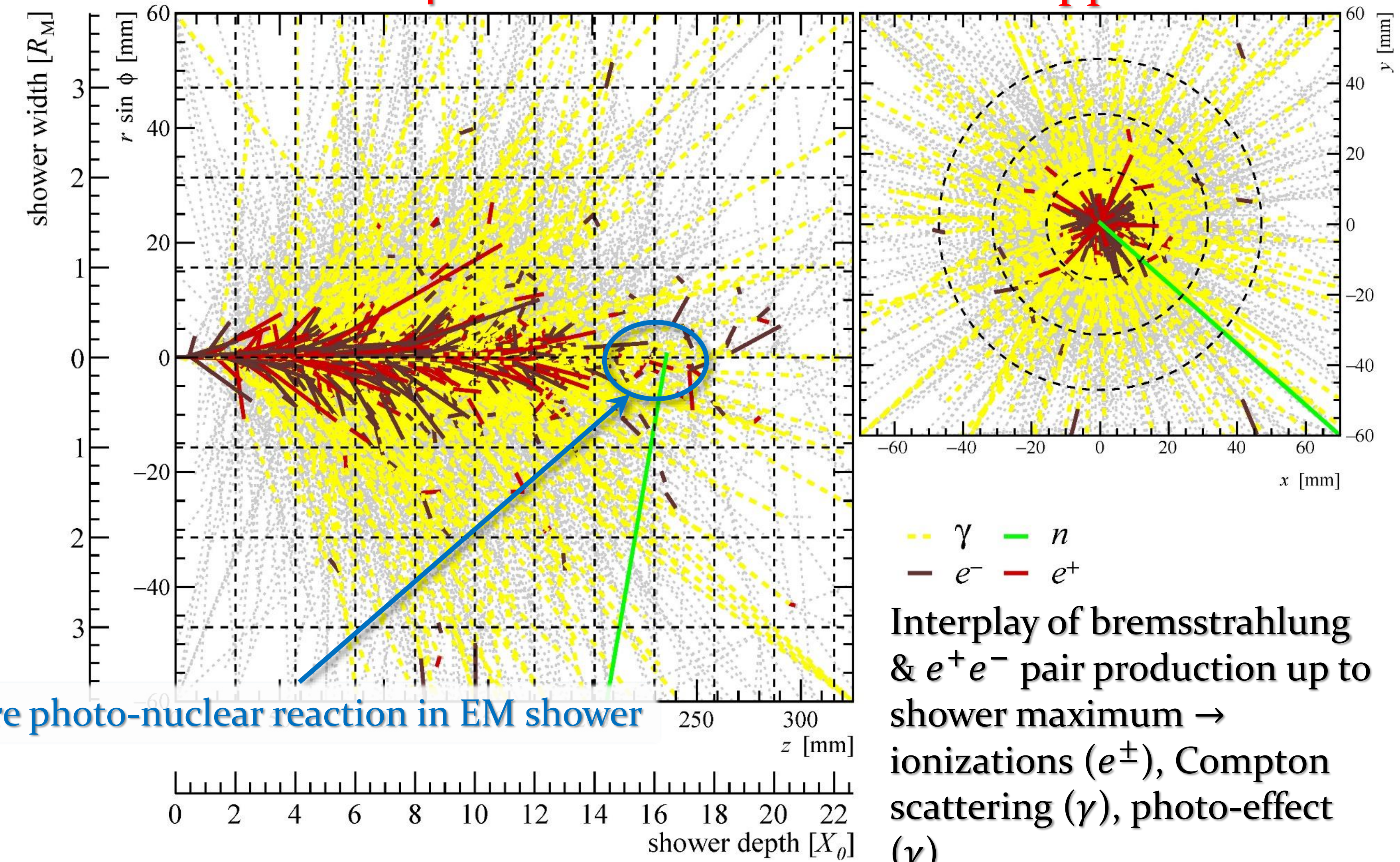


Detector Component of Interest

- ❖ ATLAS calorimeter system (similar for CMS)

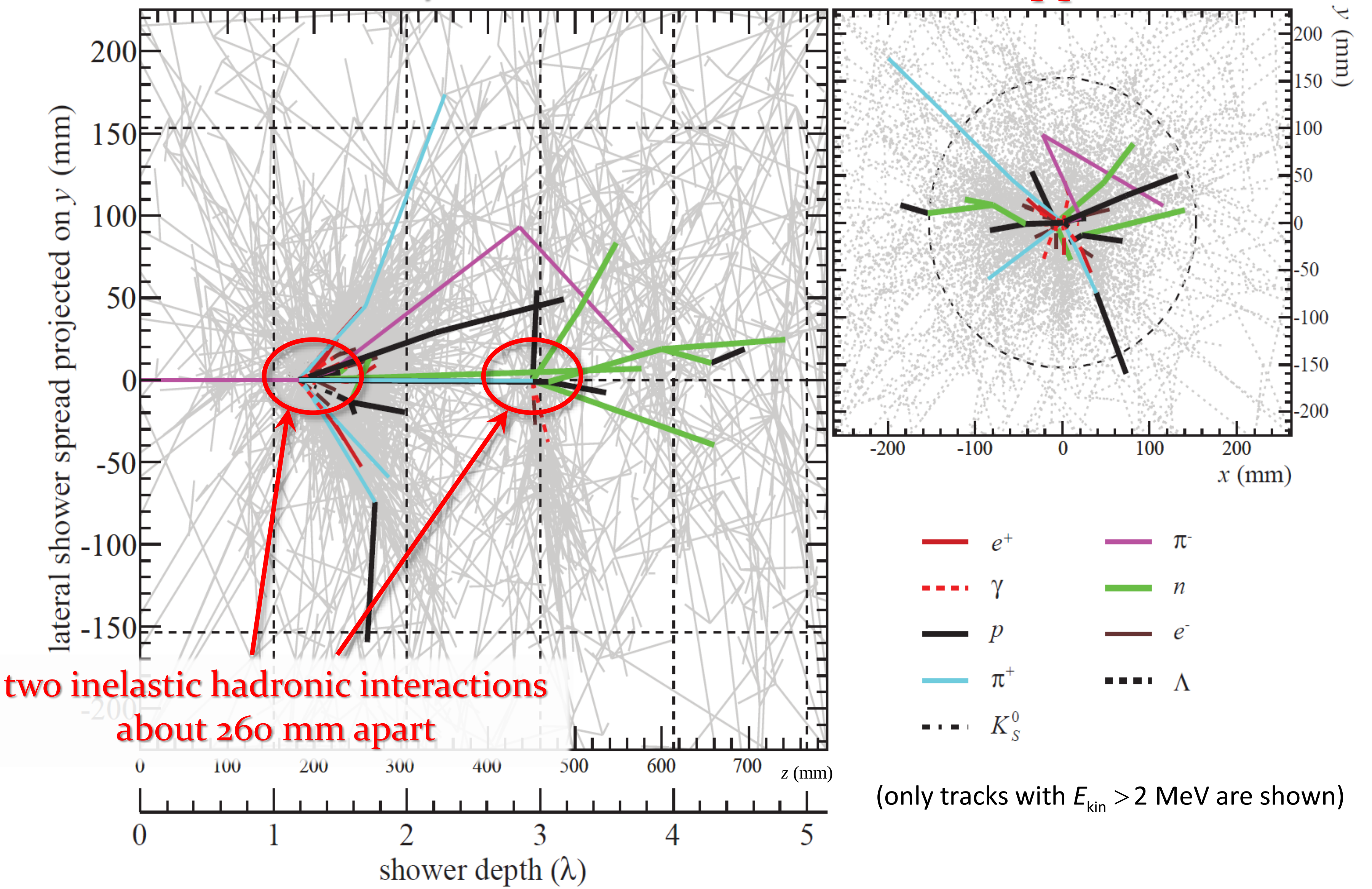


GEANT4 Simulation: 10 GeV e^- in copper



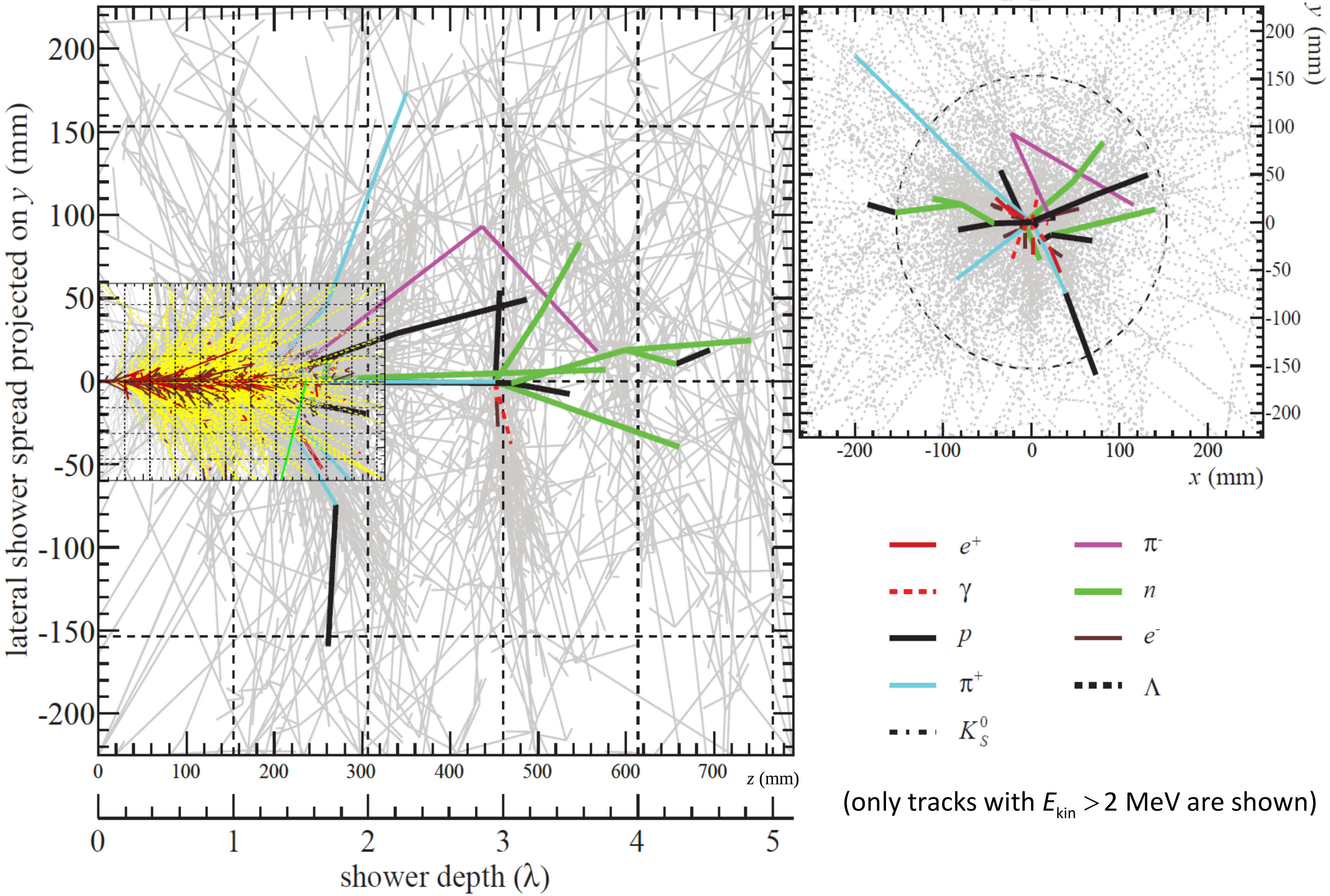
Interplay of bremsstrahlung & e^+e^- pair production up to shower maximum → ionizations ($e^\pm$), Compton scattering (γ), photo-effect (γ)

GEANT4 Simulation: 10 GeV π^+ in copper



(only tracks with $E_{kin} > 2$ MeV are shown)

GEANT4 Simulation: 10 GeV π^+ in copper

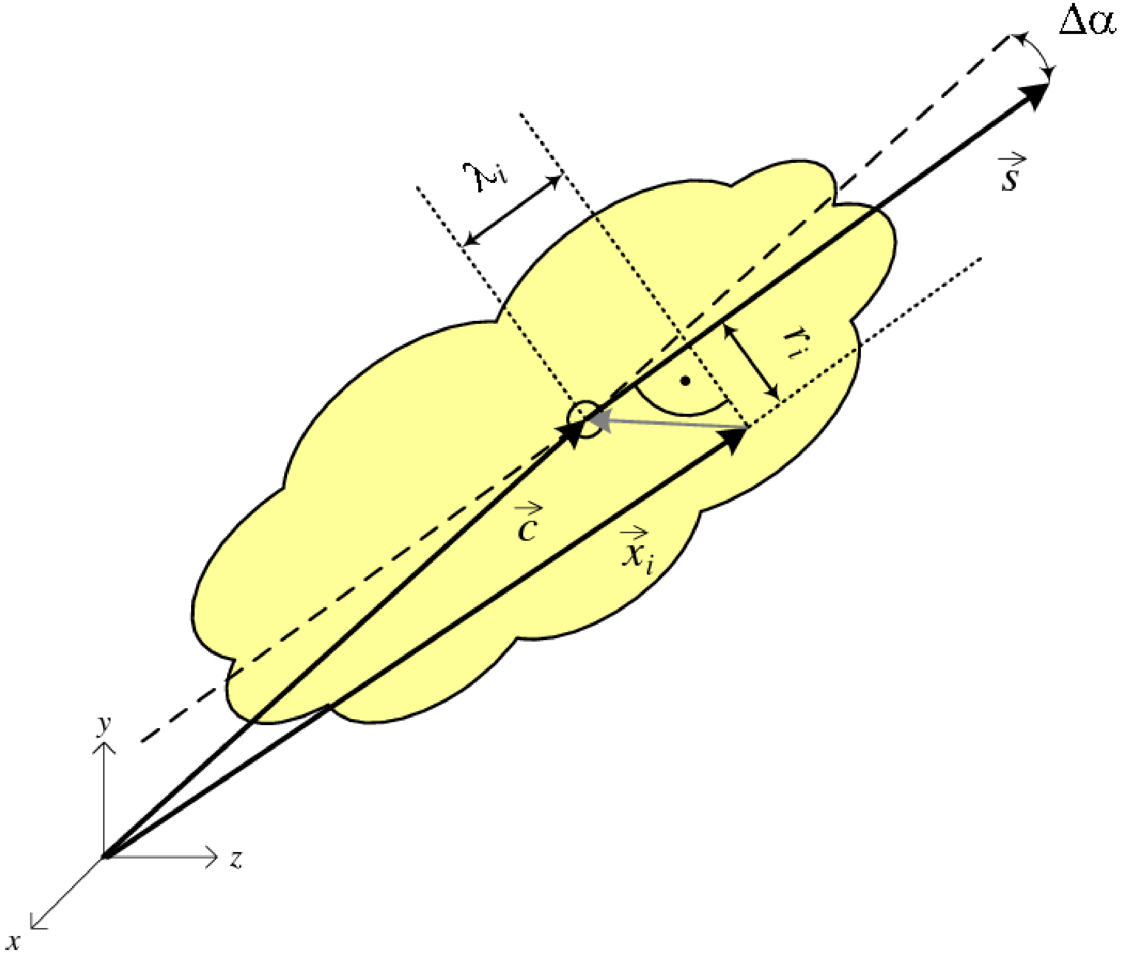


(only tracks with $E_{kin} > 2$ MeV are shown)

ATLAS Calorimeter Features

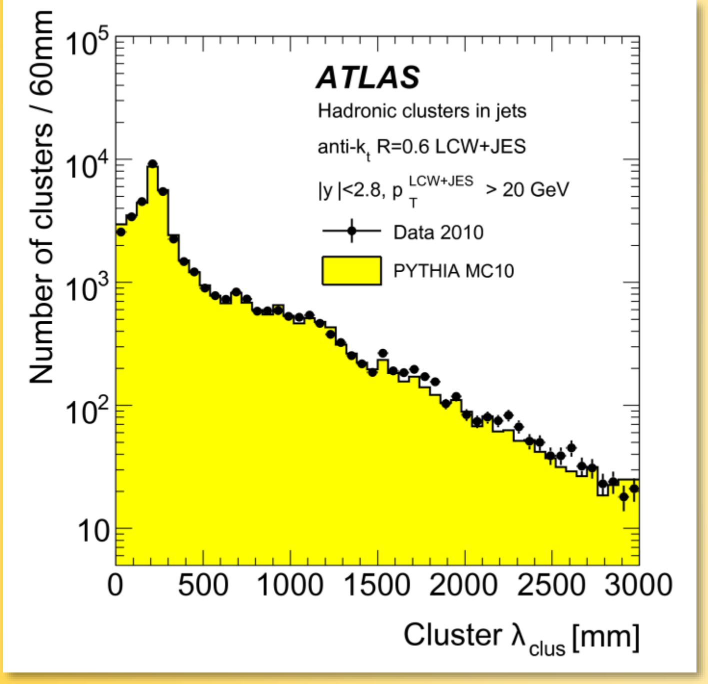
- Hardware
  - Highly granular in central region $|y| < 2.5$, sufficient granularity beyond
  - Non-compensating, hadrons generate less signal than electrons/photons depositing the same energy
- Signal extraction
  - Form three-dimensional clusters of topologically connected cell signals by following signal significance (signal-over-noise) patterns – energy blobs/**topo-cluster**
  - Algorithm features nearest-neighbor growing from a seed – collects neighbors of neighbors if signal significance of neighbor is sufficiently high
  - Applies splitting between local maxima after initial formation
  - Typically reconstructs EM shower into one topo-cluster – hadronic showers can produce > 1 topo-clusters
- Signal calibration
  - Standard LCW algorithm mitigates the non-compensation and corrects for local energy losses introduced by te clustering and losses in inactive material around the topo-cluster
  - LCW uses topo-cluster features representing the signal, the direction, the location and the shape

Geometric Topo-cluster Features



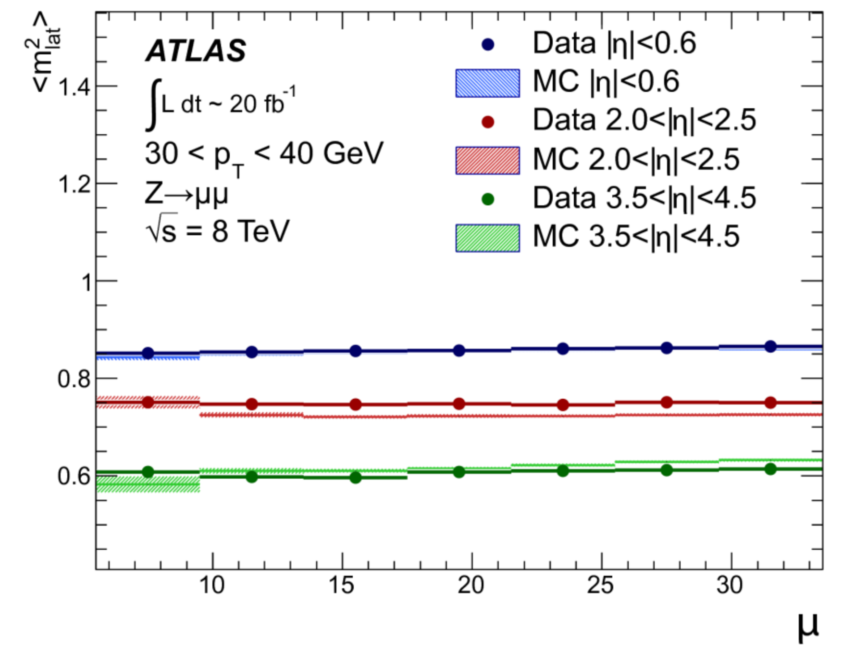
$\vec{c}$	*centre of gravity of cluster, measured from the nominal vertex* $(x = 0, y = 0, z = 0)$ *in ATLAS*
$\vec{x}_i$	*geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*
$\vec{s}$	*particle direction of flight (shower axis)*
$\Delta\alpha$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

Examples of MC Modeling



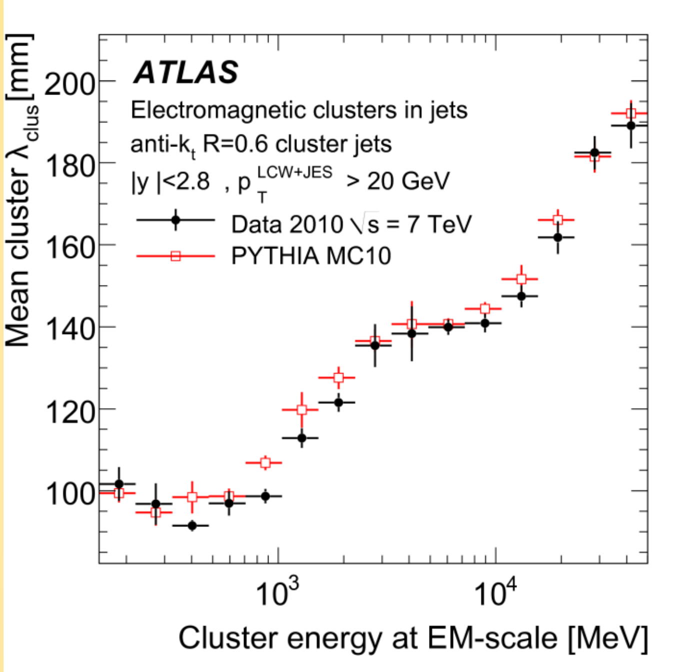
Inclusive spectrum of λ_{clus}

(topo-clusters in jets)



Pile-up dependence of $\langle m_{\text{lat}}^2 \rangle$

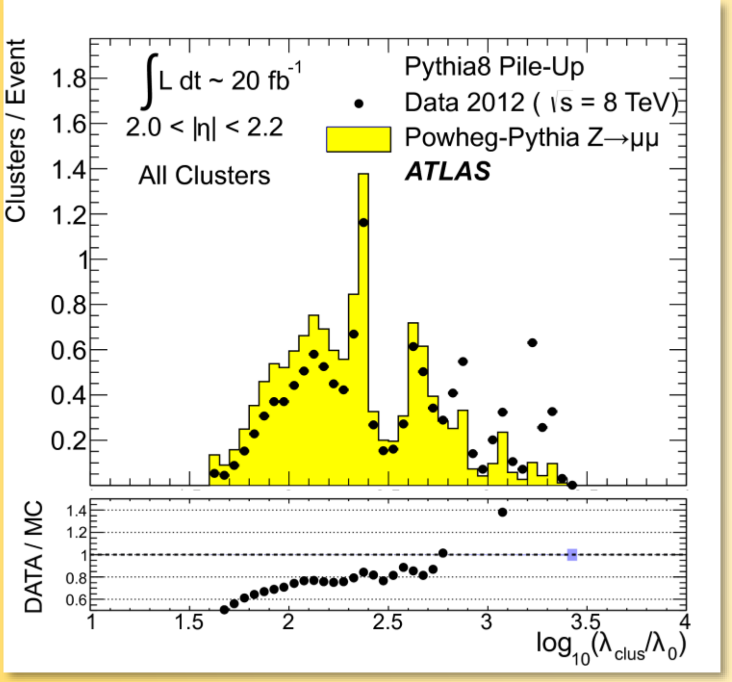
(topo-cluster in jets)



Dependence of $\langle \lambda_{\text{clus}} \rangle$ on $E_{\text{clus}}^{\text{EM}}$

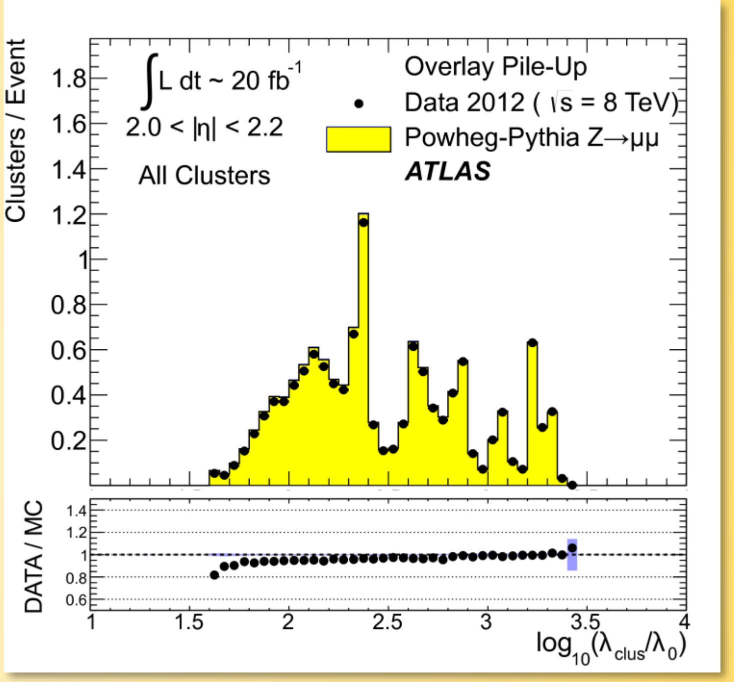
(topo-clusters in jets)

MC Modeling Problems



log(λ_{clus}) distribution of inclusive topo-cluster sample (no jet environment required)

(pile-up insufficiently modeled by MC generator & detector simulation)

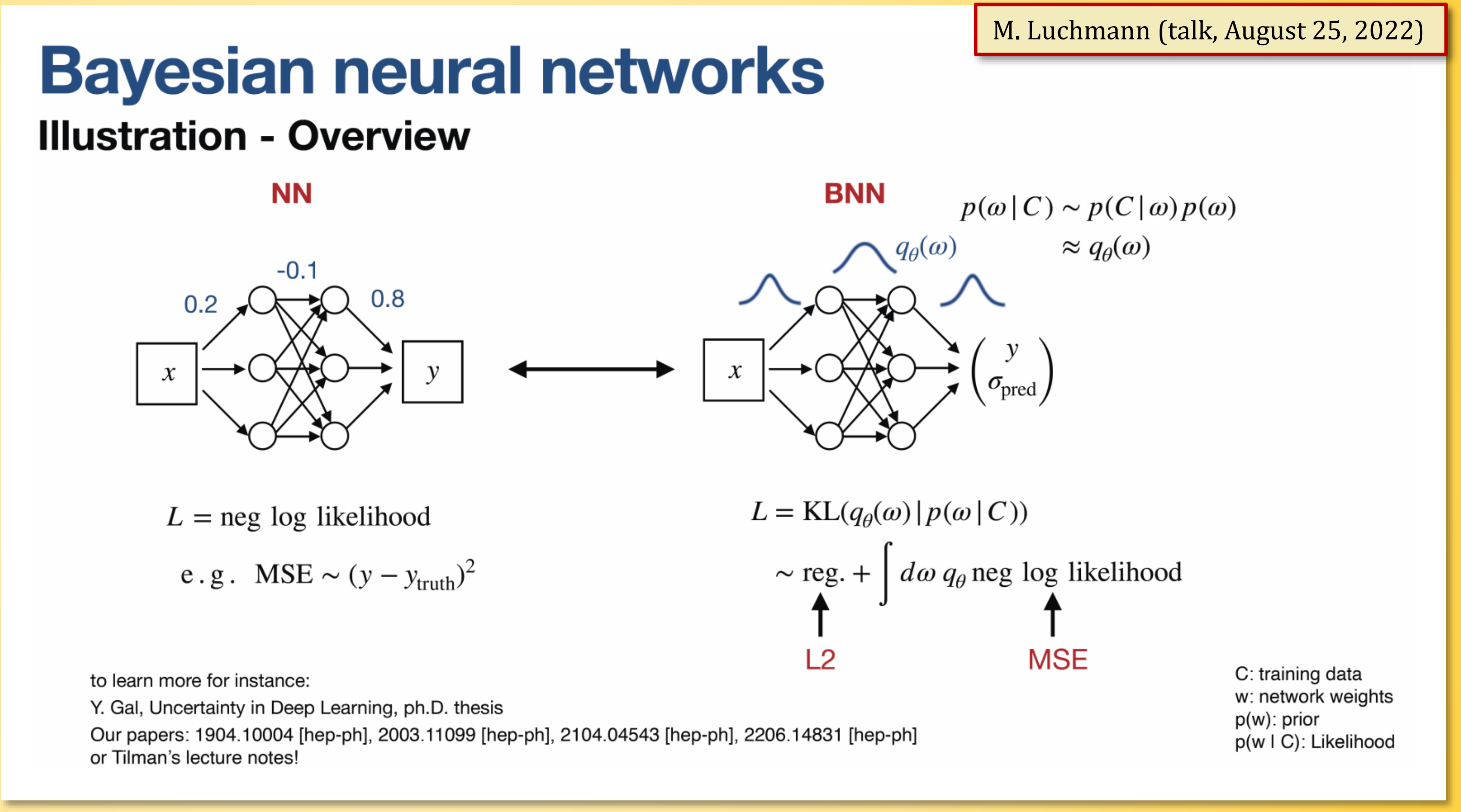


log(λ_{clus}) distribution of inclusive topo-cluster sample (no jet environment required)

(pile-up from data overlaid on hard scatter MC simulation)

Recent Considerations: BNN

- Bayesian networks
  - Principles
    - Sample of networks emulated by sampling weights from trained (Gaussian with mean $\bar{A}(w)$ and width $\sigma_{\mathrm{stoch}}(w)$) weight distributions $q_\theta(w)$ instead of training fixed weights w
    - **Contributes uncertainties** due to sampling ($\sigma_{\mathrm{stoch}}(w)$) from each network and due to training multiple networks simultaneously ($\sigma_{\mathrm{pred}}(w)$) $\Rightarrow$ **calibration model uncertainties**



Inputs from, and discussions with, **P.A. Delsart** & **Ana Peixoto** (both LPSC Grenoble), **Chris Delitzsch** (University of Dortmund) and a lot of advice, technical (implementation) help and code from **T. Plehn** & **M. Luchmann** (both University of Heidelberg)

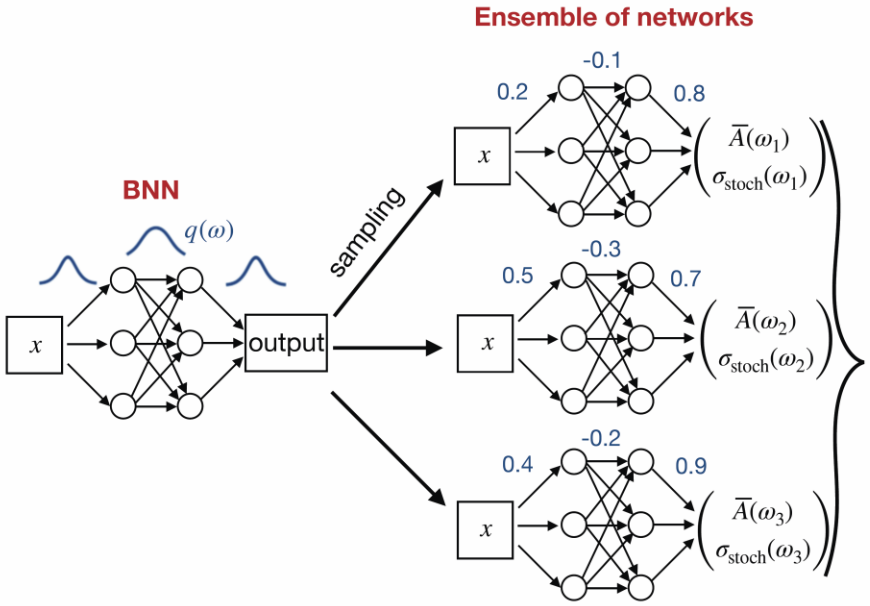
Recent Considerations: BNN

- Bayesian networks
  - Principles
    - Sample of networks emulated by sampling weights from trained (Gaussian with mean $\bar{A}(w)$ and width $\sigma_{\text{stoch}}(w)$) weight distributions $q_\theta(w)$ instead of training fixed weights w
    - **Contributes uncertainties** due to sampling ($\sigma_{\text{stoch}}(w)$) from each network and due to training multiple networks simultaneously ($\sigma_{\text{pred}}(w)$) $\Rightarrow$ **calibration model uncertainties**

M. Luchmann (talk, August 25, 2022)

Bayesian neural networks

Illustration - BNN as an ensemble of ordinary networks



Output

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^{N} \bar{A}(w_i)$$

$$\sigma^2_{\text{stoch}} = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{\text{stoch}}(w_i)$$

$$\sigma^2_{\text{pred}} = \frac{1}{N} \sum_{i=1}^{N} (\langle A \rangle - \bar{A}(w_i))^2$$

Inputs from, and discussions with, **P.A. Delsart** & **Ana Peixoto** (both LPSC Grenoble), **Chris Delitzsch** (University of Dortmund) and a lot of advice, technical (implementation) help and code from **T. Plehn** & **M. Luchmann** (both University of Heidelberg)