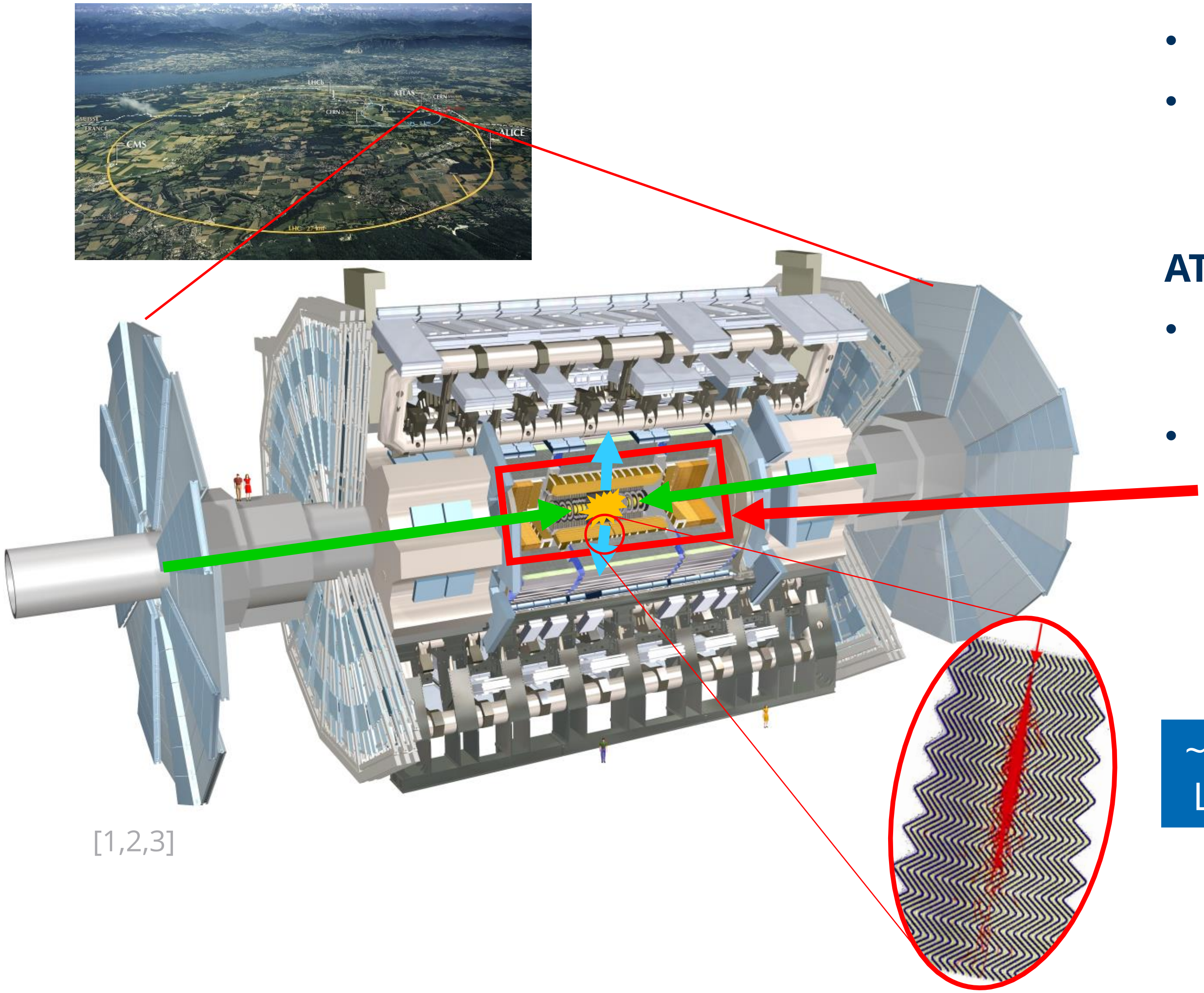
Machine Learning for Real-Time Processing of ATLAS Liquid Argon Calorimeter Signals with FPGAs

Anne-Sophie Berthold

PUNCH Meeting
18th September 2023

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

The ATLAS Detector at the LHC



[1,2,3]

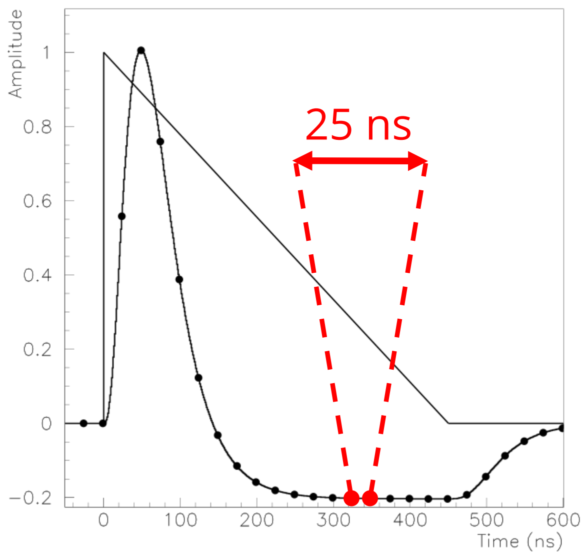
Large Hadron Collider (LHC):

- Proton bunches collide with **25 ns** spacing (40 MHz)
- 2029: Start of **High Luminosity LHC** (HL-LHC) with up to ~7 x nominal luminosity

ATLAS Detector → Phase-II upgrade (2026-2028)

- From ~20 collisions to up to ~200 collisions per bunch crossing (BC) → pileup & trigger rate increases
- Readout electronics of **Liquid-Argon (LAr) calorimeter** needs to be improved

~182 000
LAr cells



[4]

LAr Calorimeter Readout

FPGA – Field Programmable Gate Array

[7]

- Real-time signal processing
- Installation of **556 high-performance FPGAs**

Optimal Filtering Algorithm (OF)

- calculates deposited energy per cell

$$y_t = \sum_{n=0}^{M-1} x_{t-n} \cdot a_n$$

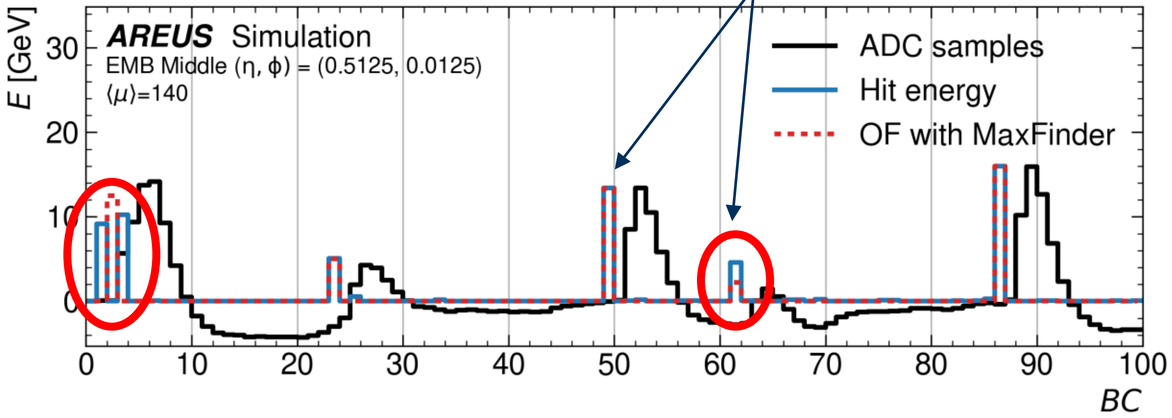
y_t ... OF output for bunch crossing (BC) t
$\vec{x}$ ... input ADC samples
a_n ... OF coefficients
M ... OF filter depth

Trigger

Event Readout

- Trigger system applies additional maximum finder

- Good in energy resolution **but**
- Weak in reconstruction of overlapping signals



➢ Development of Neural Networks to improve energy reconstruction

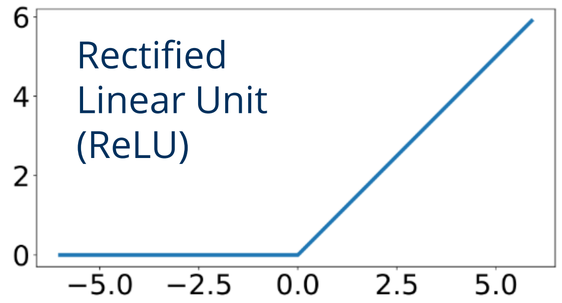
Convolutional Neural Networks (CNNs)

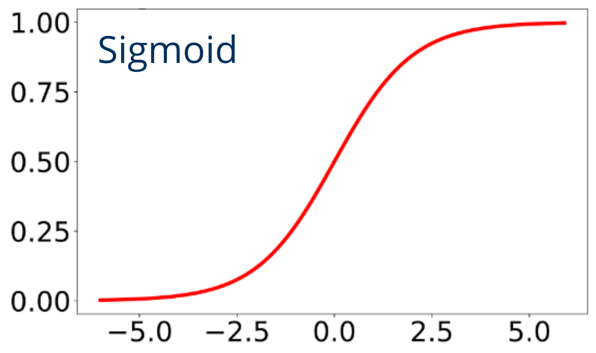
- Convolutional operation with certain **kernel** size
- **Activation function** gives opportunity to classify, weight, cut

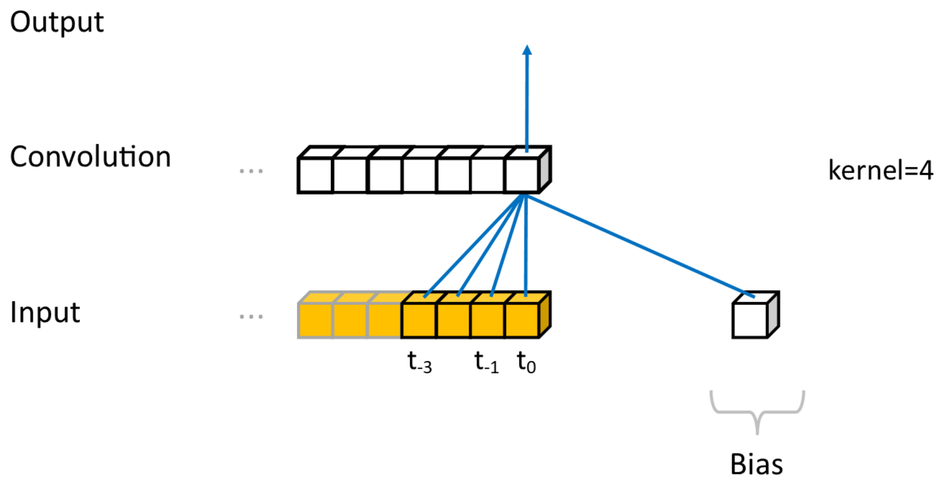
$$y_t = A\left(\sum_{i=0}^{n} x_i \cdot w_i + b\right)$$

y_t … node output for time t
$\vec{x}$ … input samples
$\vec{w}$ … weights
b … bias term
A … activation function

→ **similar to OF**

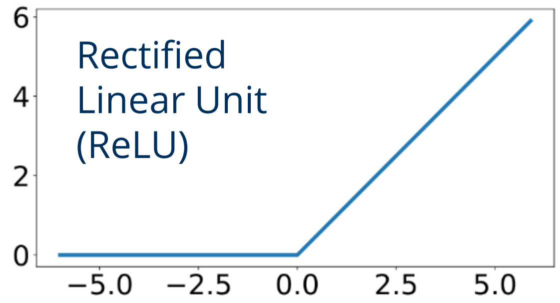


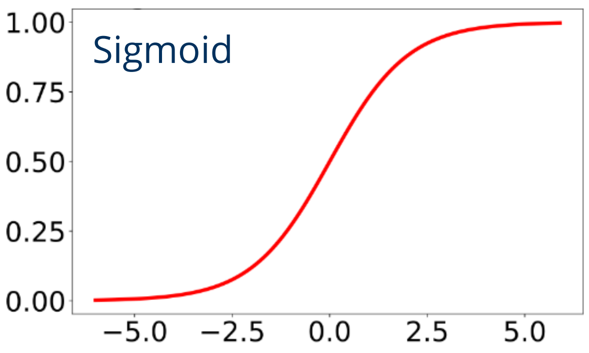


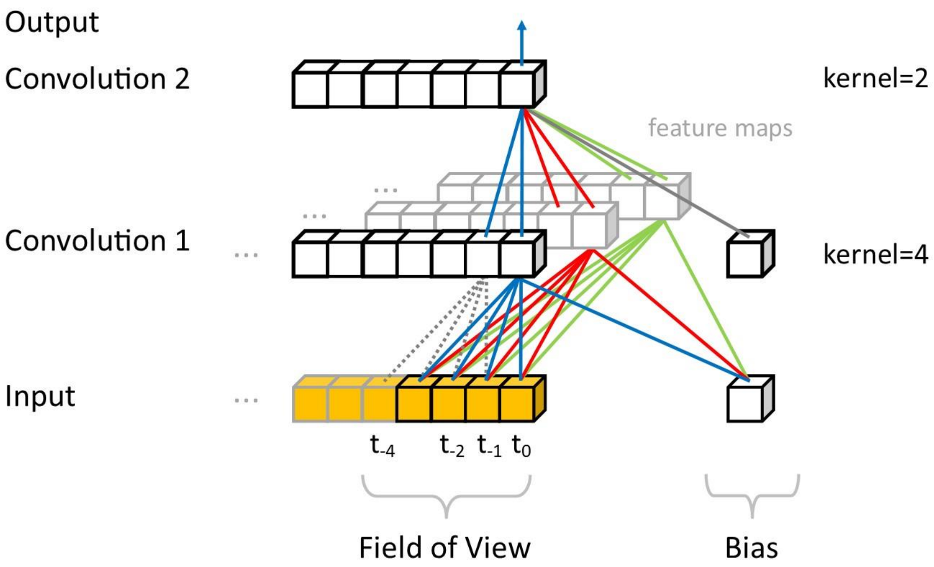


Convolutional Neural Networks (CNNs)

- Convolutional operation with certain **kernel** size
- **Activation function** gives opportunity to classify, weight, cut
- **Feature maps** focus on different properties
- **Training** minimizes difference between output and target

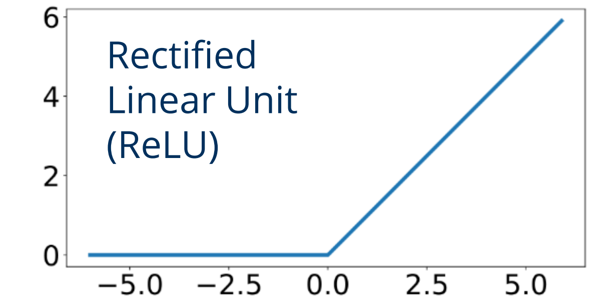


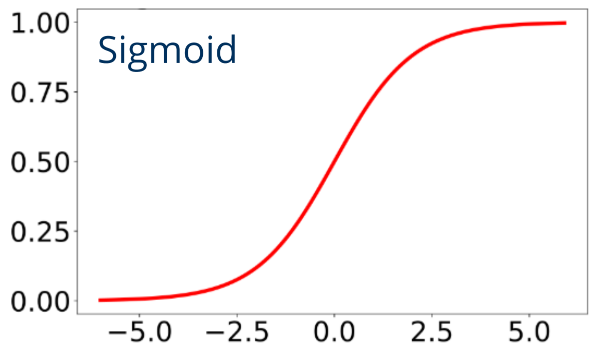


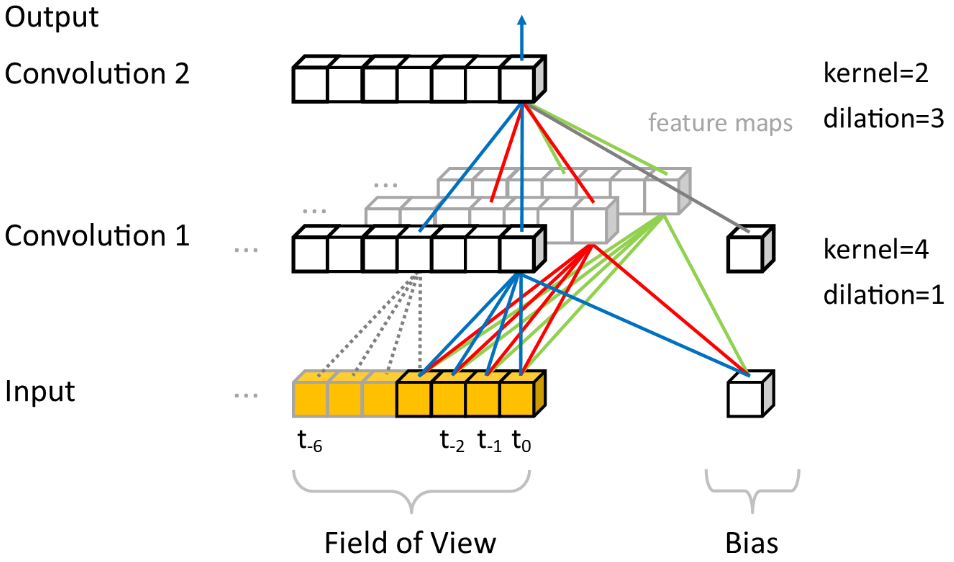


Convolutional Neural Networks (CNNs)

- Convolutional operation with certain **kernel** size
- **Activation function** gives opportunity to classify, weight, cut
- **Feature maps** focus on different properties
- **Training** minimizes difference between output and target
- **Dilation** varies field of view (FoV) without increasing parameters
- **Keep parameters low** (≈ 100 /≈ 400) and FoV realistic (≤ 24) due to FPGA implementation







CNN Training Workflow

[6,7]

- Programming environment: **Python / Keras / Tensorflow**
- Framework on **CERN GitLab**

- Different approaches for general architecture
- For each: application of hyperparameter search with **Keras Tuner** [a]
  - ***Hyperband* Algorithm** for faster scanning of hyperparameter search space [b]
  - Limit to *N* parameters requires parameterization of search space

- Once architecture found, repeat training $\sim 100x$ to estimate **reproducibility** of CNN
- Use loss as first estimate to choose „best" weights (simply mean absolute error)
- Secondly: look into performance measures

- Training/Evaluation data from ***AREUS* simulation** (C++ - CERN)
  - Easy to produce high amount of data
  - Need to emulate all possible scenarios and balance ratio of importance in learning data
  - Tuning preliminary as final influence i.e. on particle reconstruction not yet studied in detail

[a]: https://keras.io/keras_tuner/ [b]: https://arxiv.org/abs/1603.06560

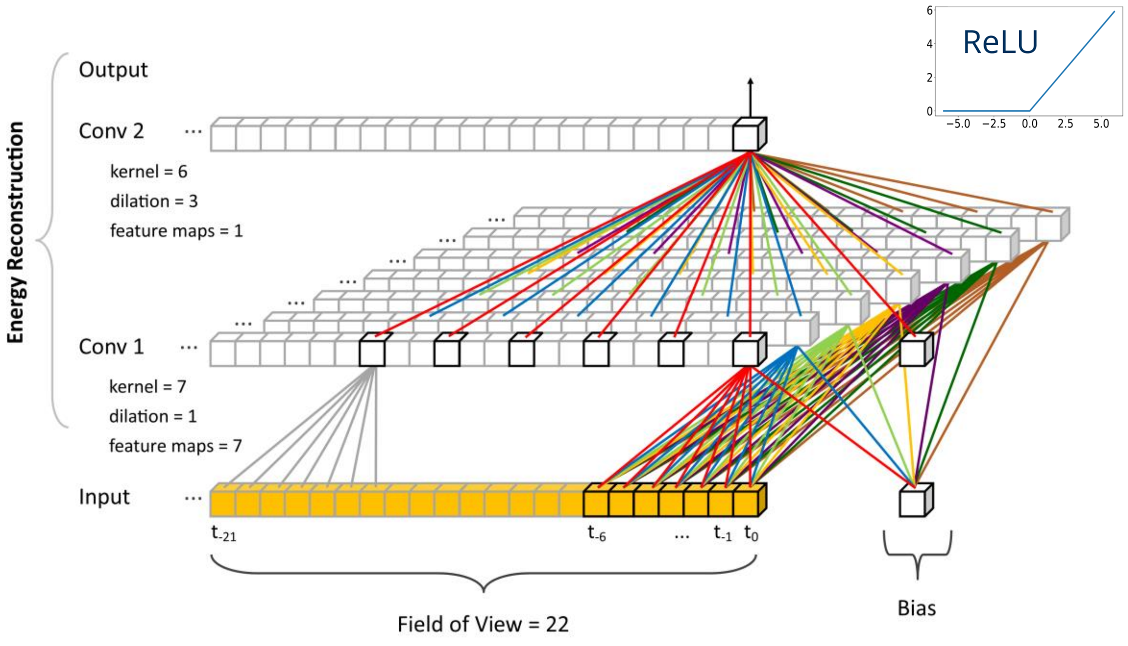
CNNs for LAr Readout: Two Recent Approaches

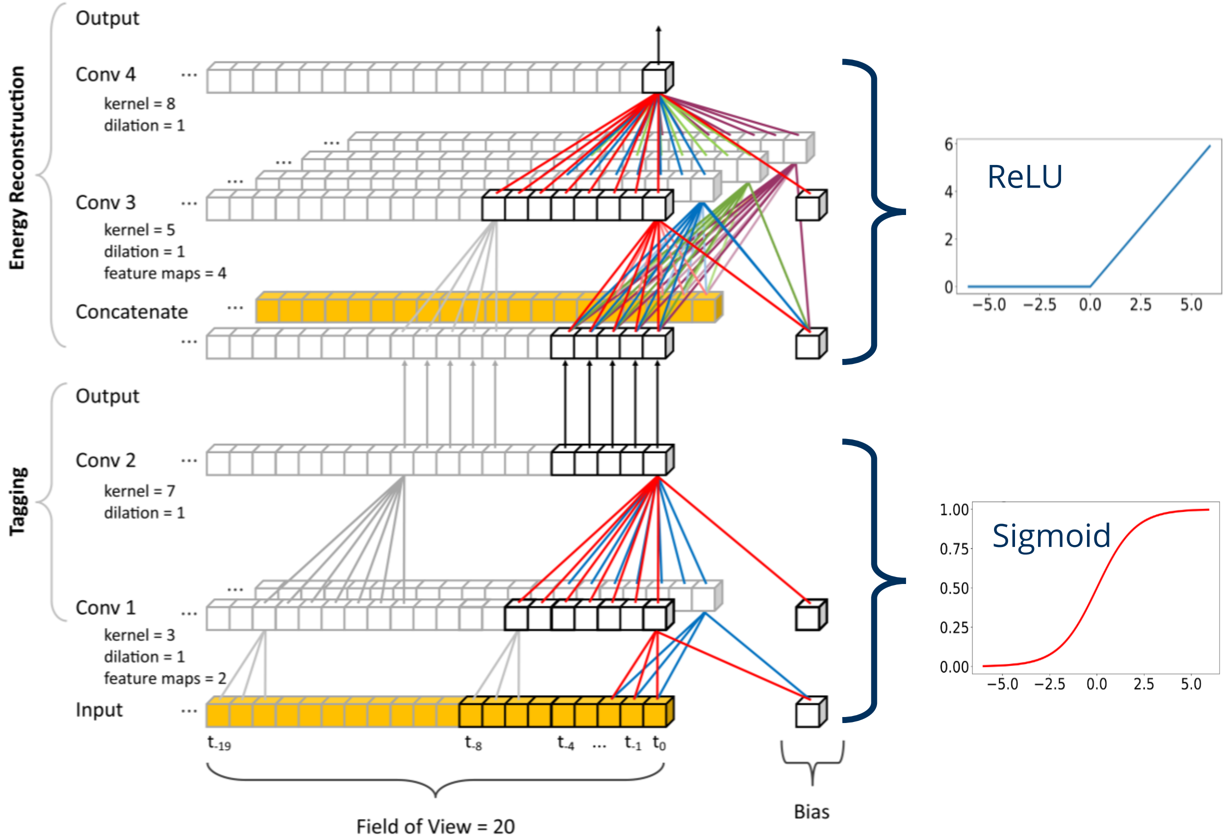
Plain 2-layered CNN (2CNN)

- Dilation enables larger Field of View (FoV)
- ReLU activation functions
- Output: reconstructed energy

4-layered CNN with Tagging (4TCNN)

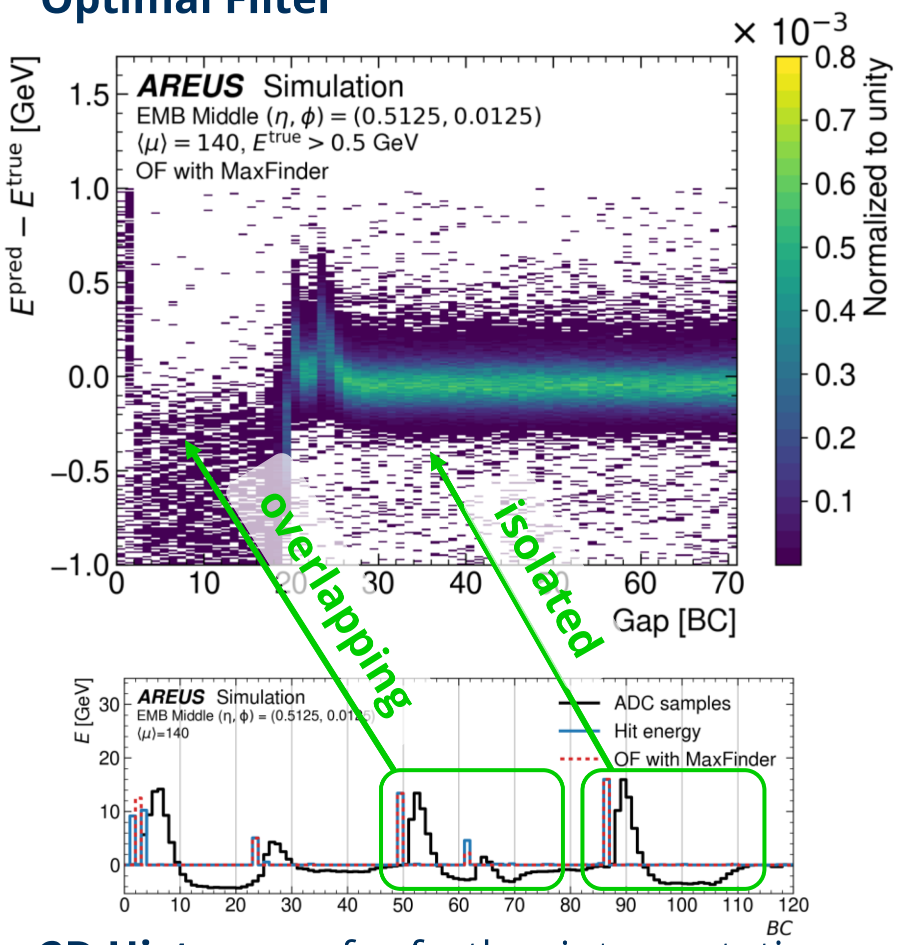
- Sigmoid and ReLU activation functions
- Intermediate output tags signal overlaps
- Output: reconstructed energy



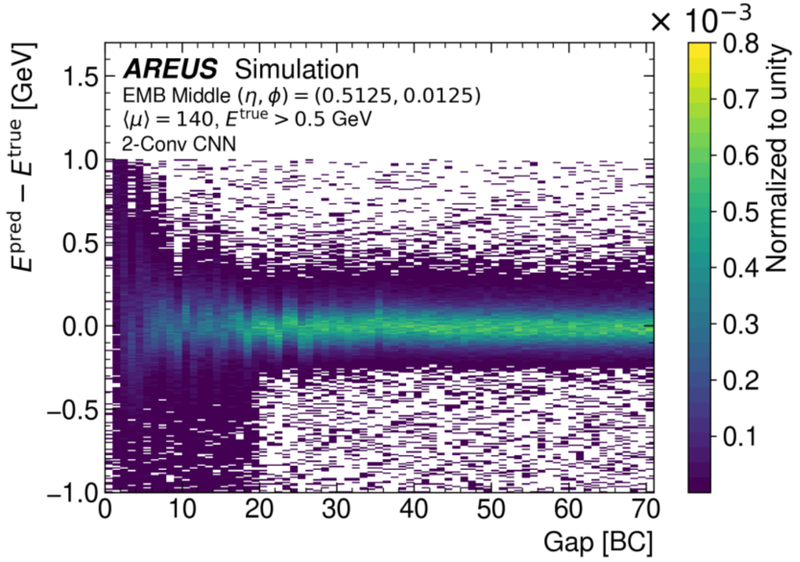


Performance Measure: Energy Reconstruction as Function of Gap

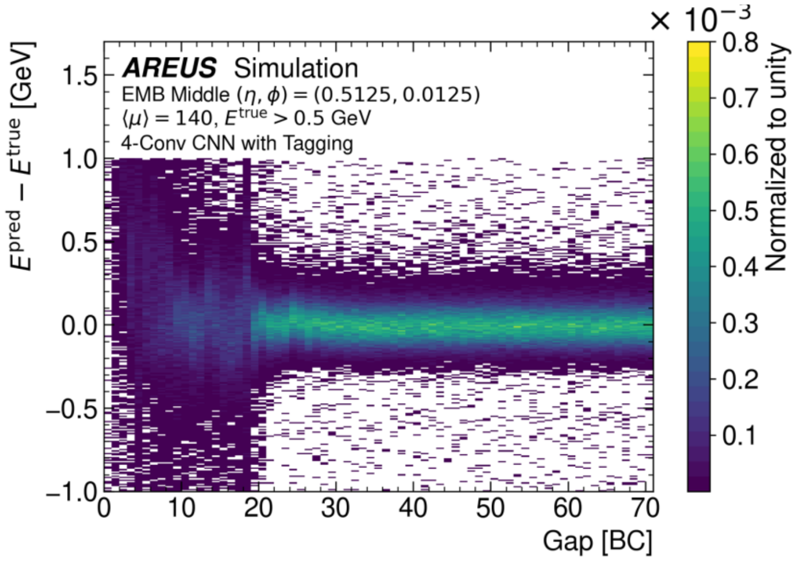
Optimal Filter



2-Conv CNN



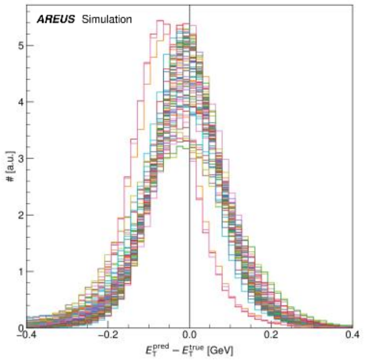
4-Conv CNN with Tagging

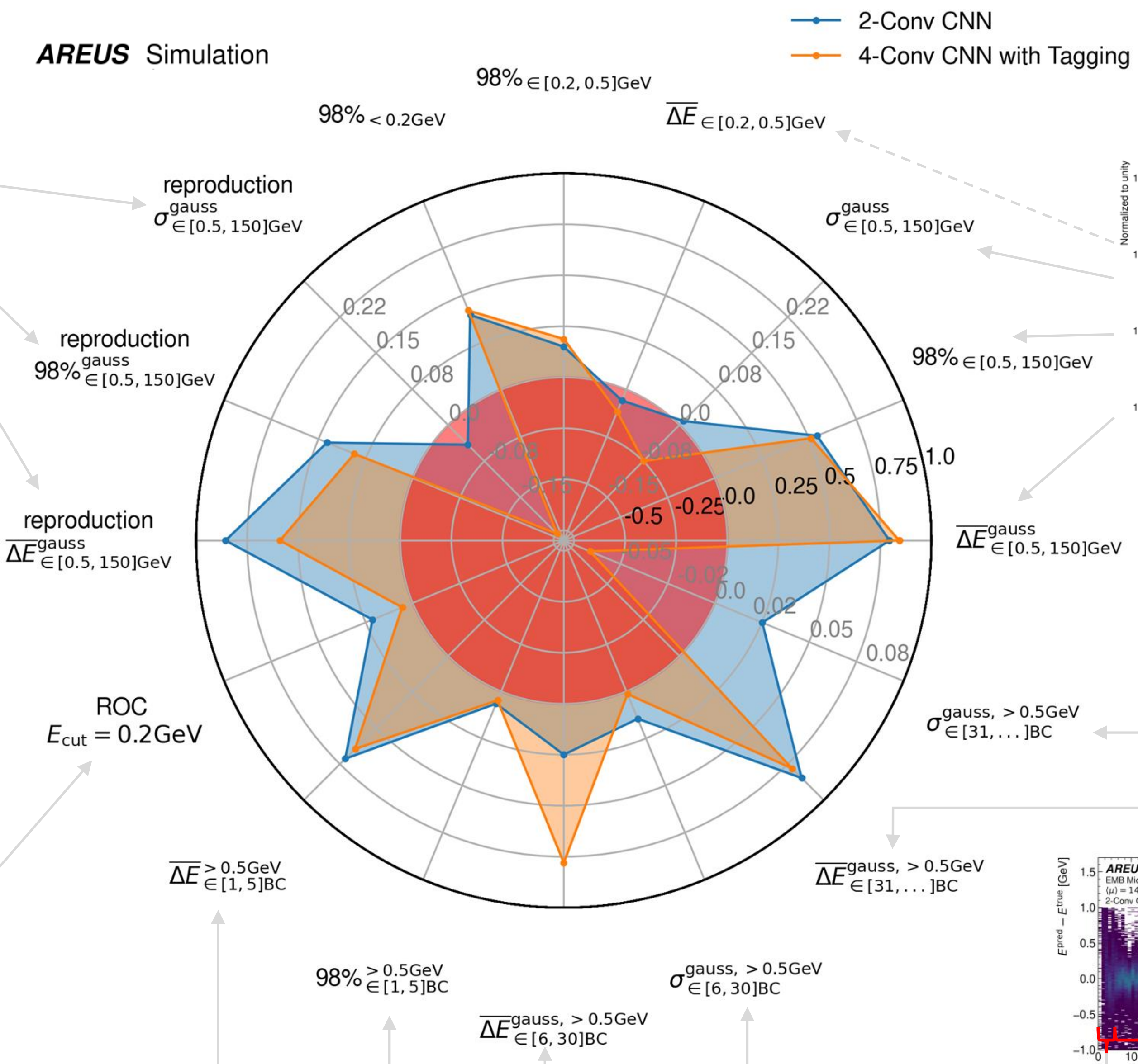


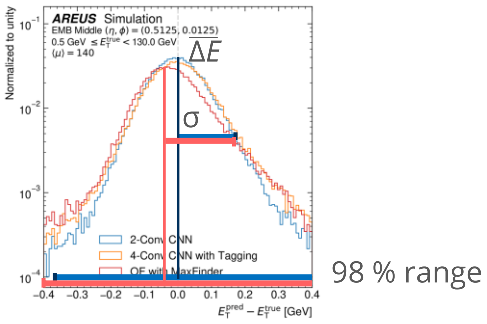
2D Histogram for further interpretation

➢ OF struggles with overlapping pulses (gap < 25 BC), CNNs show improvement

Performance Evaluation - Combining all: Star Plot



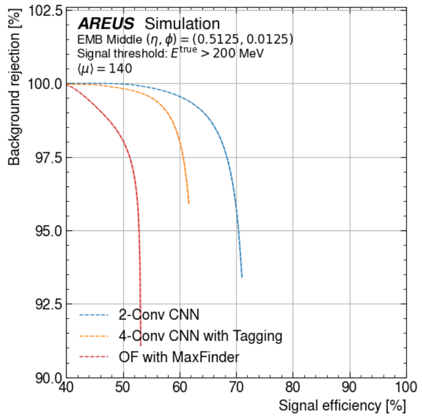




Score indicates CNN performance:

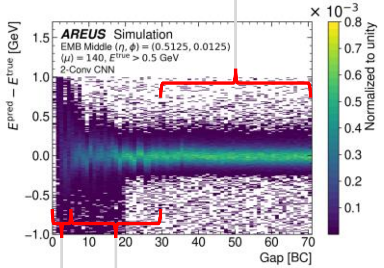
$$sc(X) = 1 - \frac{X\_ANN}{X\_OF}$$

- **Red circle:** (=0) OFMax yield
- **Outer circle:** (=1) best yield
- Inside circle (<0): worse than OF



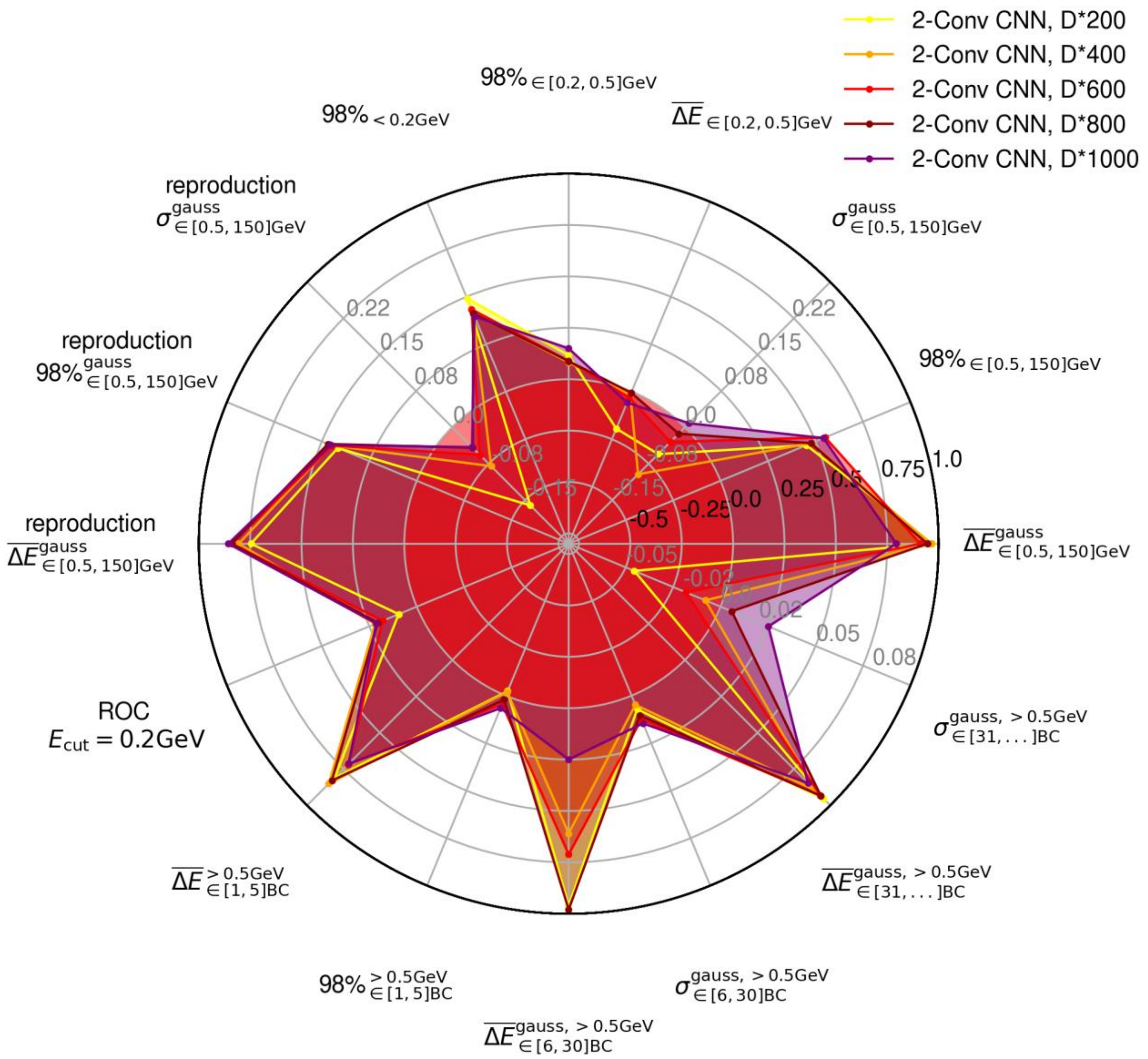
✓ Performance overview

x No replacement for other plots as details might be hidden



TECHNISCHE UNIVERSITÄT DRESDEN

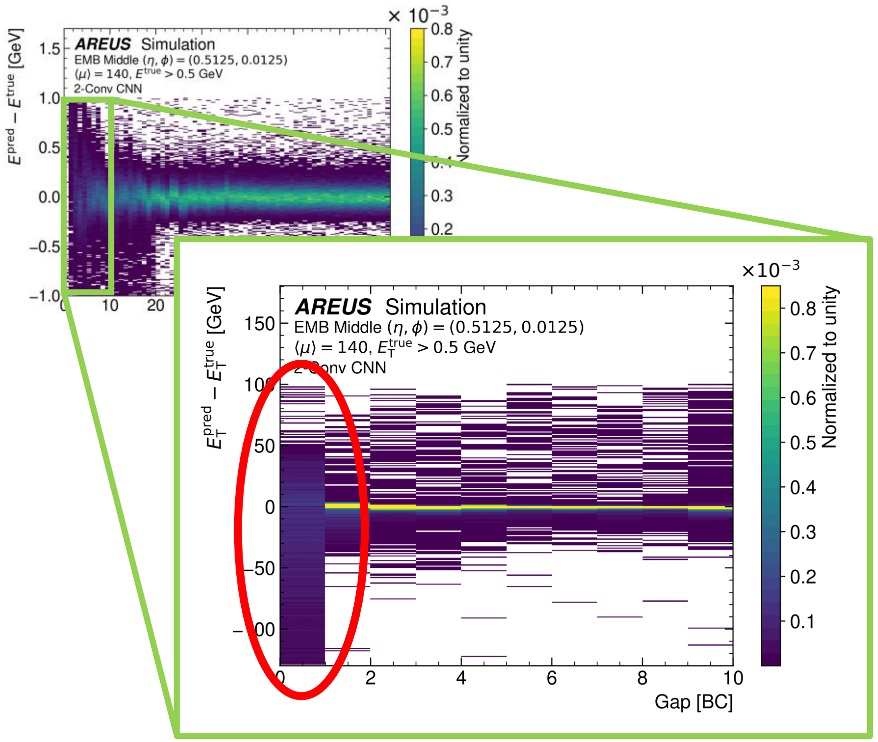
Varying Size of Training Dataset N

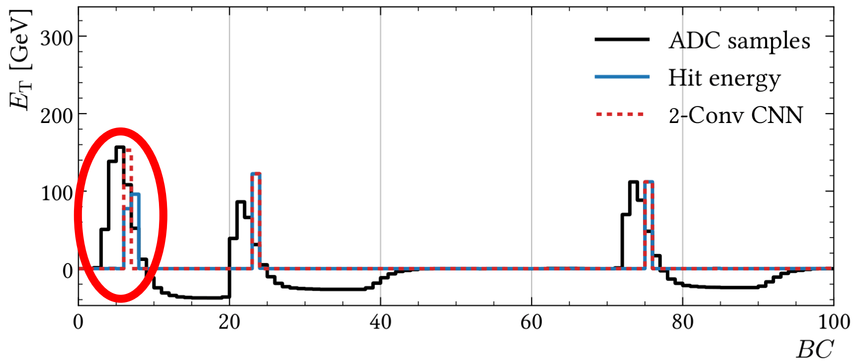


Equally enhance all 6 training sub-datasets:

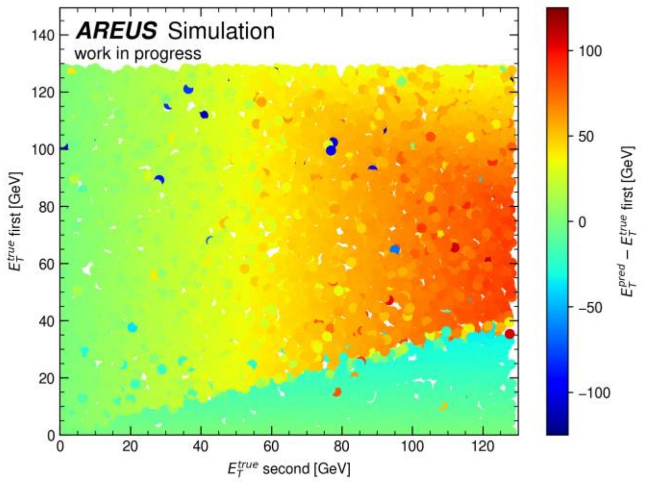
- [200, 400, 600, 800, 1000]*10,000 BC for each scenario
- Some scores not affected
- For others: at least 600*10,000 BC for each

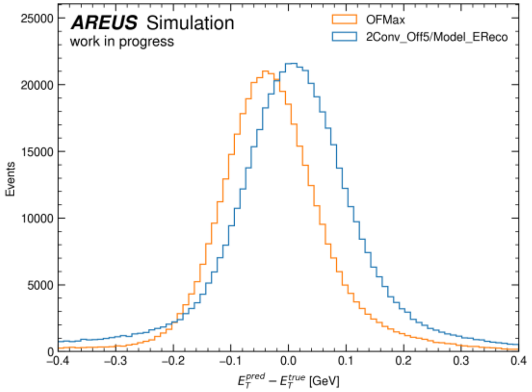
Improving gap=0





Initial training

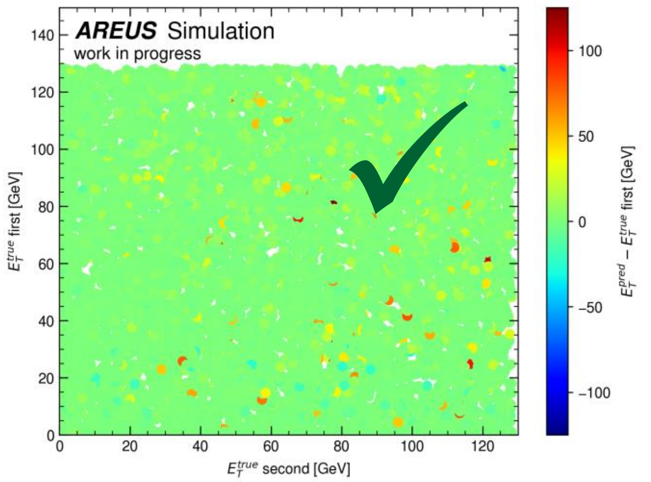


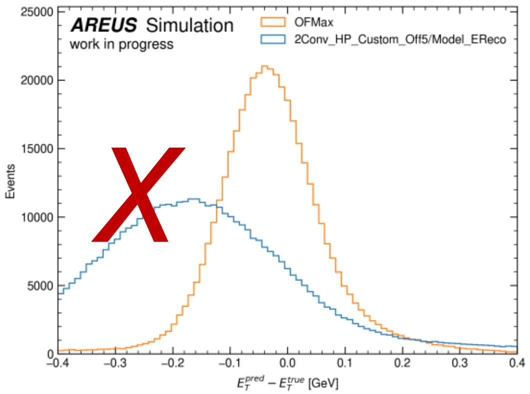


Zero gap not resolvable by CNNs

→ Add another dataset to training dataset which holds 0 gaps

Training with Zero Gap





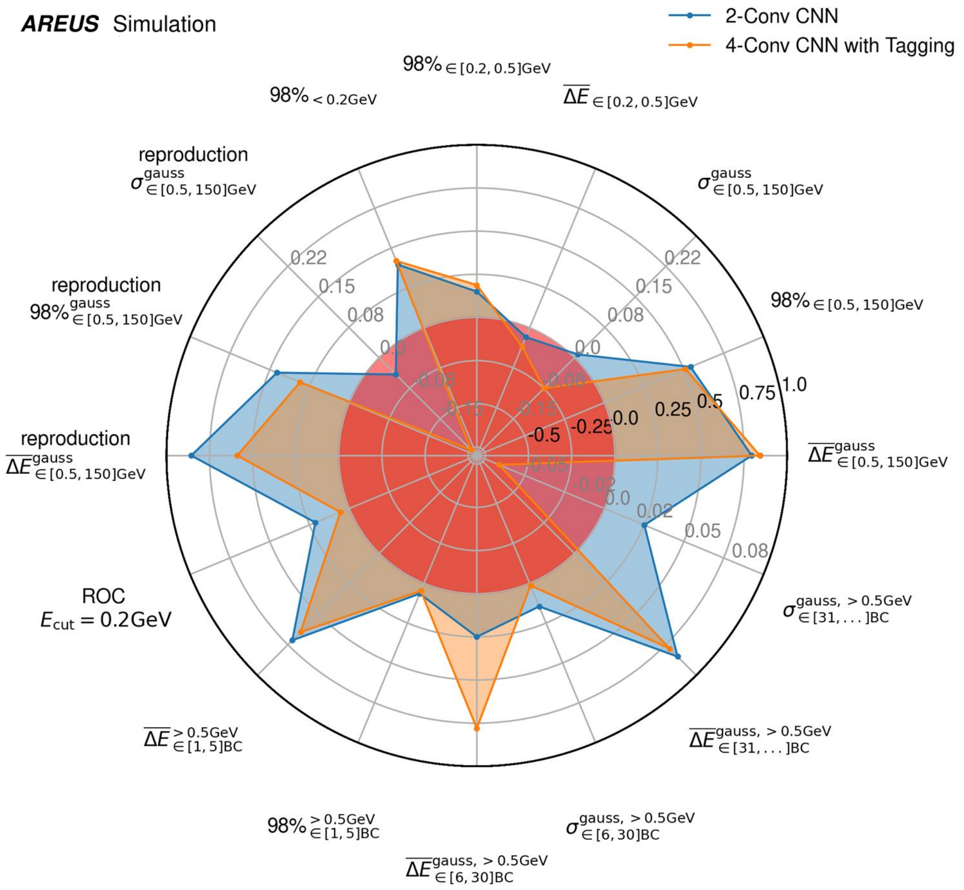
Zero gap behavior improved but energy resolution drops

→ optimize ratio of 0 gap occurrence in training data e.g. via additional parameter for hyperparameter search

Summary and Outlook

Summary

- ANNs are able to replace Optimal Filtering algorithm
- FPGA resource requirements regarding latency and bandwidth can be satisfied
- Different performance requirements must be met or weighed against each other
- Visualisation in star plot gives overview



Outlook

- Further improvements by applying quantization aware training and more CNN parameters
- Study for influence of energy reconstruction by ANNs for full event reconstruction
- Further tests on FPGA hardware ongoing

Thank you for your attention!

Sources I

Slide 2:

[1] URL: https://static1.bmbfcluster.de/3/4/3/8_ef6a5eef8f44963/3438meg_22ce2885dae52af.jpg.

[2] Joao Pequenao. *Computer generated image of the whole ATLAS detector.* CERN. Mar. 27, 2008. URL: https://cds.cern.ch/record/1095924 (visited on 05/10/2023).

[3] Karl Jakobs. Lecture Material. CERN. 2015. URL: https://www.particles.uni-freiburg.de/dateien/vorlesungsdateien/particledetectors/kap8

[4] ATLAS Collaboration. *Monitoring and data quality assessment of the ATLAS liquid argon calorimeter.* CERN. May 13, 2014. URL: https://cds.cern.ch/record/1701107 (visited on 05/24/2023).

Slide 3:

[5] Intel. *Stratix 10 FPGA*. URL: https://newsroom.intel.com/editorials/intels-stratix-10-fpga-supporting-smart-connected-revolution (visited on 04/18/2021).

Sources II

Slides 7:

[6] *Keras Logo*. URL: https://keras.io/ (visited on 05/25/2023)

[7] *Tensorflow Logo*. URL: https://www.vectorlogo1.zone/logos/tensorflow/index.html (visited on 05/25/2023)

Papers related to these slides:

- Georges Aad et al. *Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS Lar Calorimeters*. In: Computing and Software for Big Science 5.1 (Oct. 2021) DOI: 10.1007/s41781-021-00066-y. URL: https://doi.org/10.1007/s41781-021-00066-y.
- Georges Aad et al. *Firmware implementation of a recurrent neural network for the computation of the energy deposited in the liquid argon calorimeter of the ATLAS experiment*. 2023. DOI: 10.48550/ARXIV.2302.07555. URL: https://doi.org/10.48550/arXiv.2302.07555.

Used Training Data

Training data contain sequences with 6 different scenarios

Furthermore: artificial loss enhancement for signal hits >1 GeV by factor=30 (keras parameter <sample_weight>)

No.	Bunch Filling	Signal Gap	Energy Range	Bunch Length	N x 10,000 BCs
	Determines LHC filling pattern – succession of pp collisions	Occurrance of high-E signal hits	high: up to 80% low: up to 10% of possible energy range	leads to variation in timing of incoming pulse	Relative number in training dataset
1)	each filled	const, gap=45	high	5 cm	1/6
2)	each filled	const, gap=45	low	5 cm	1/6
3)	each filled	None	only pile-up	5 cm	1/6
4)	each filled	random, gaussian: mu=30, sigma=10	high	5 cm	1/6
5)	each filled	random, gaussian: mu=30, sigma=10	low	5 cm	1/6
6)	each filled	random, gaussian: mu=30, sigma=10	high	50 cm	1/6

total fraction of low energetic hits enlarged to push energy resolution in this range

overlapping signals

Larger variance of pulse timing

Sequence Comparison

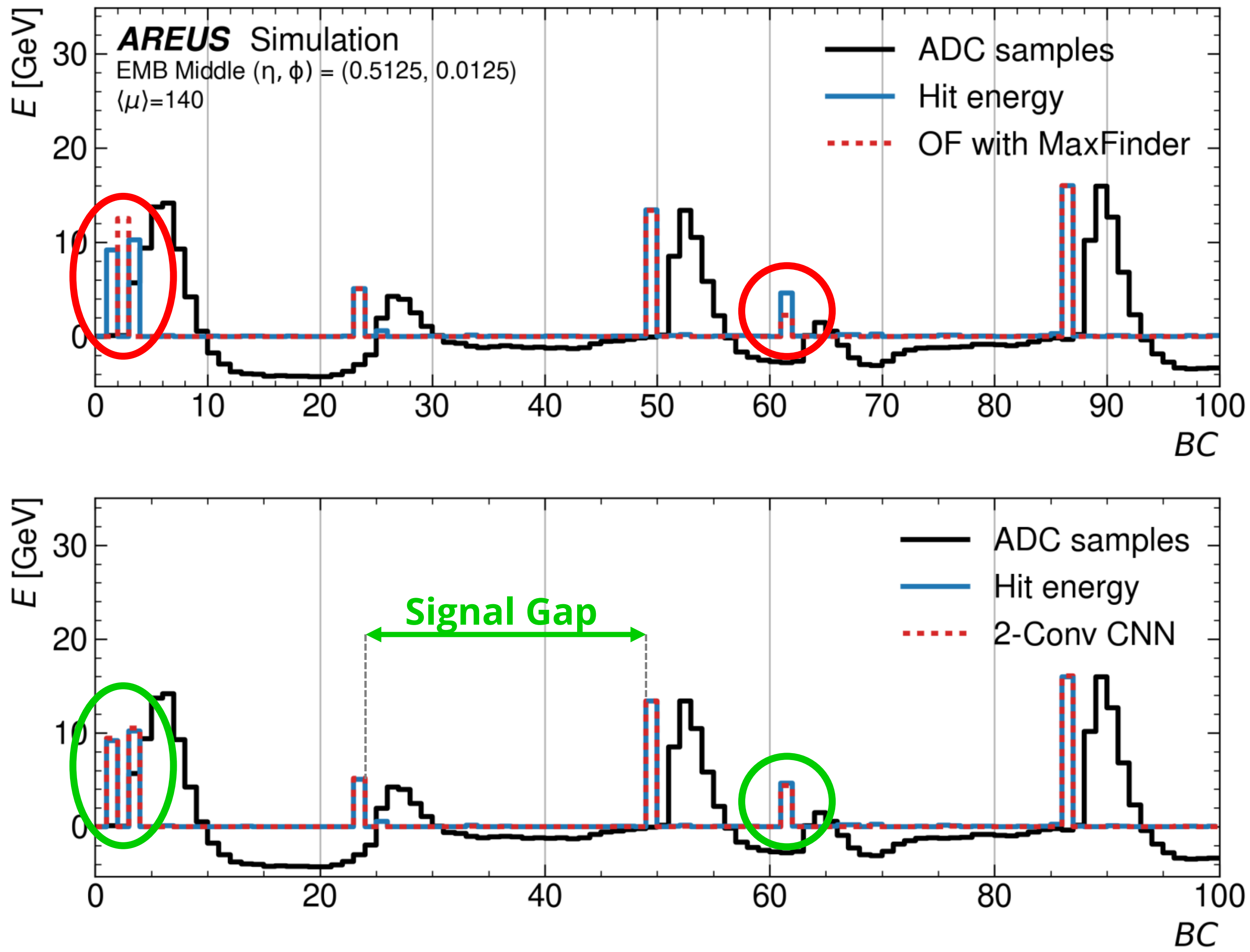
Example sequence for few events

I.e.: scenario where Optimal Filter struggles:

- Close signals cannot be resolved
- Signals within undershoot underestimated

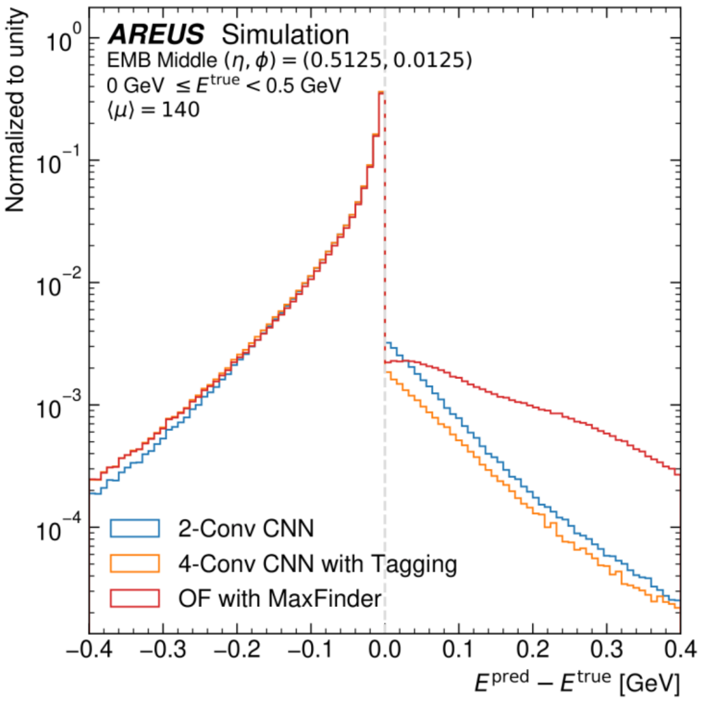
ANN:

- Optimized to reconstruct overlapping signals

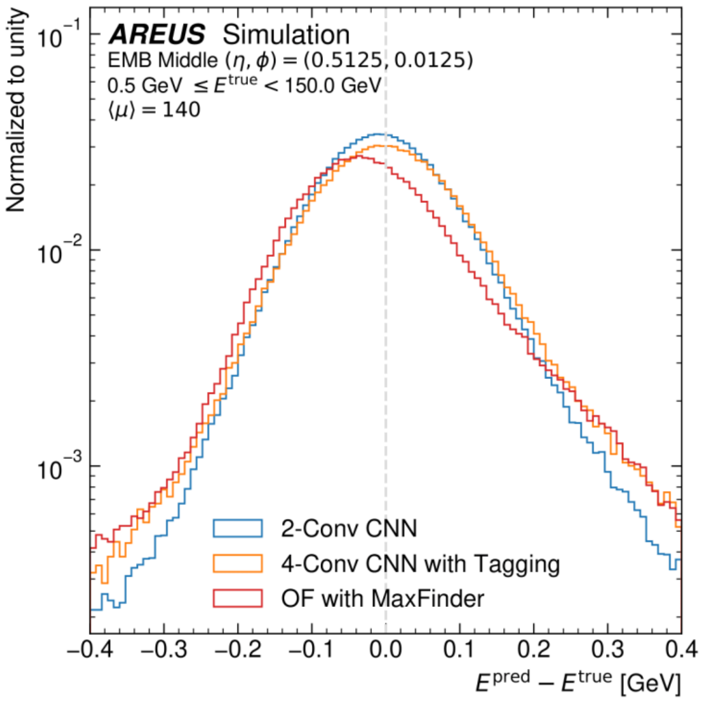


First Performance Measure: Energy Resolution

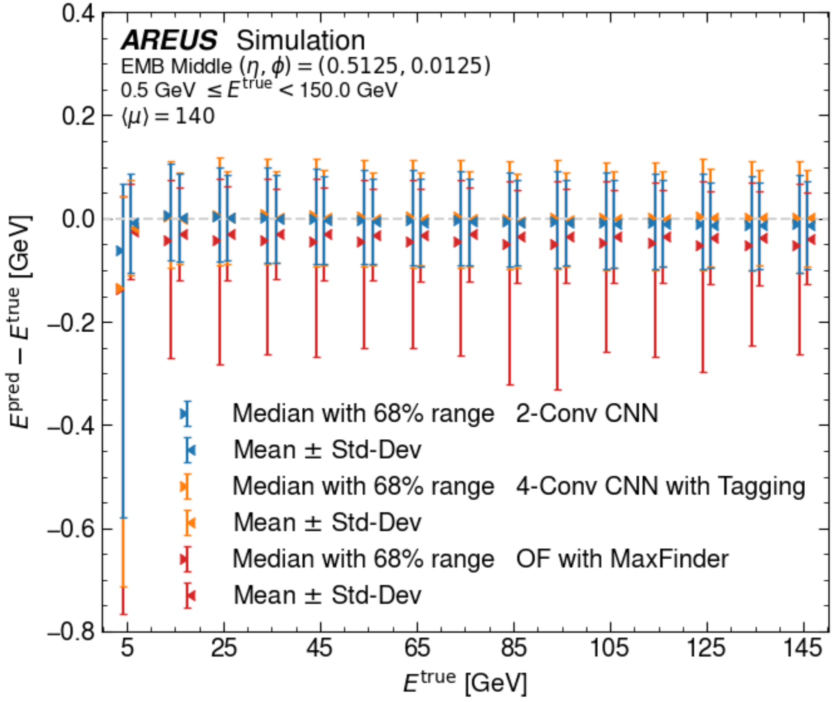
$E^{true} < 0.5$ GeV



$E^{true} > 0.5$ GeV



Median and Mean over E^{true}



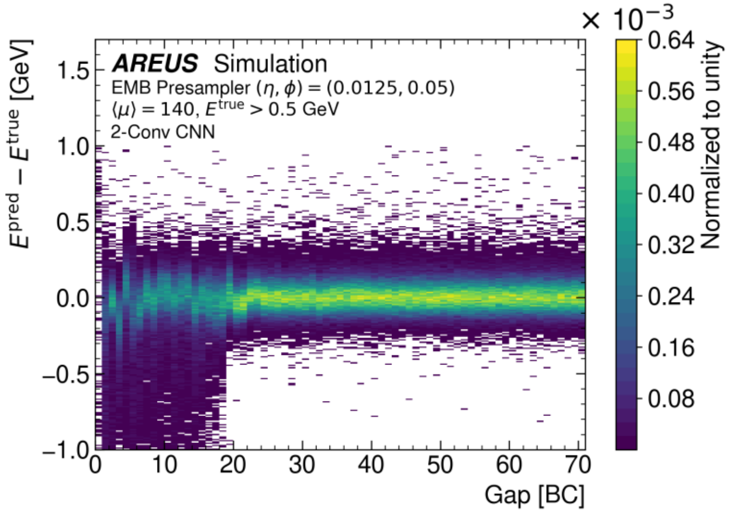
1D Histogram for overall energy resolution

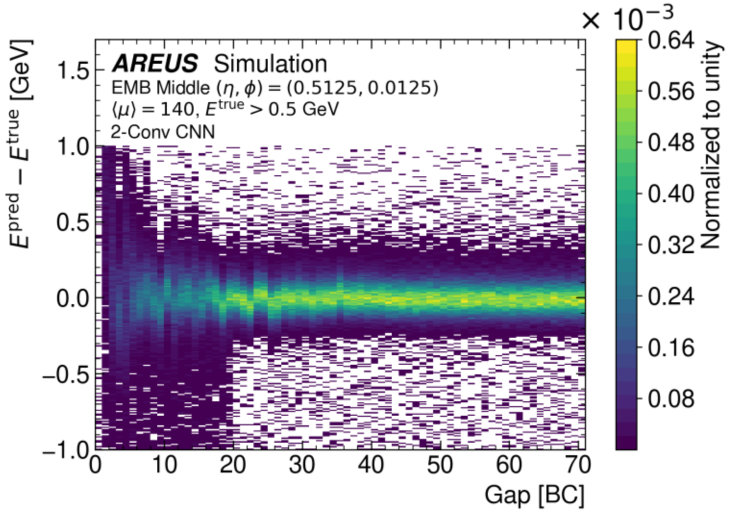
Optimal Filter: larger deviation spread in low energy region, negative bias in high energy region

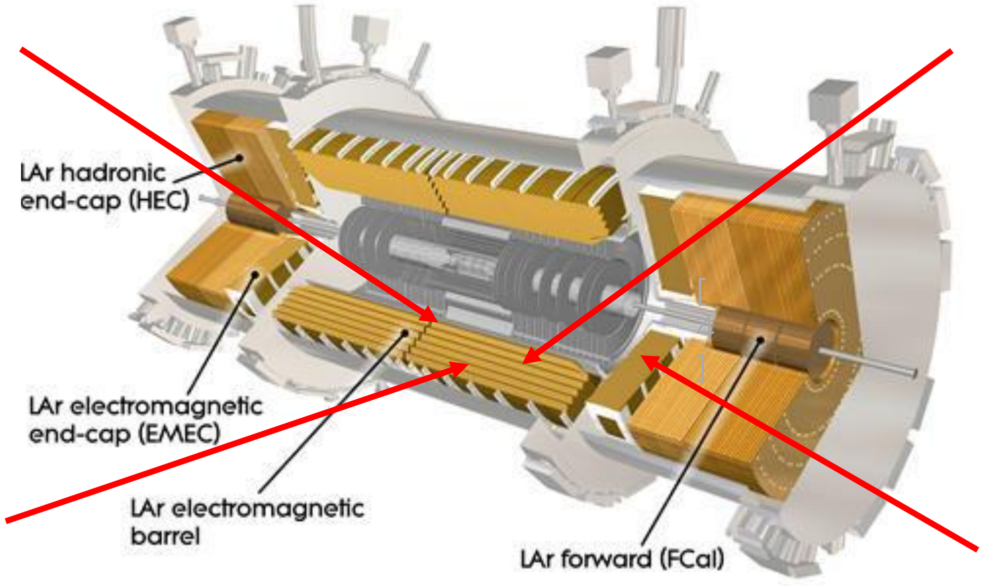
CNNs: improvement in energy resolution, more even and centered distribution

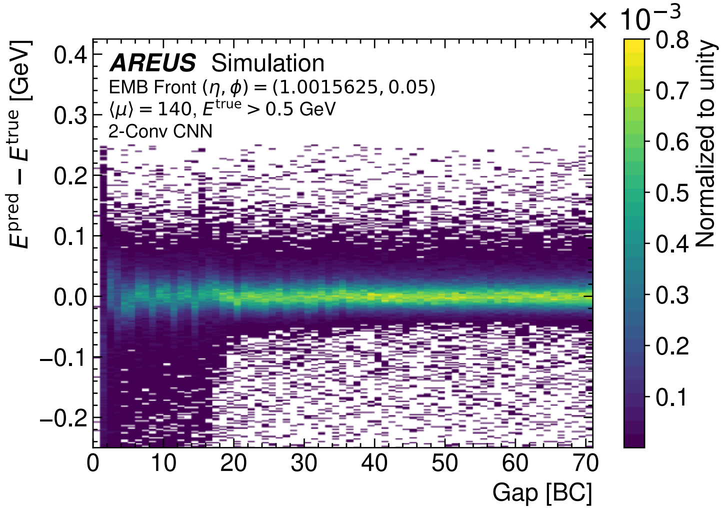
Performance stable within large energy range

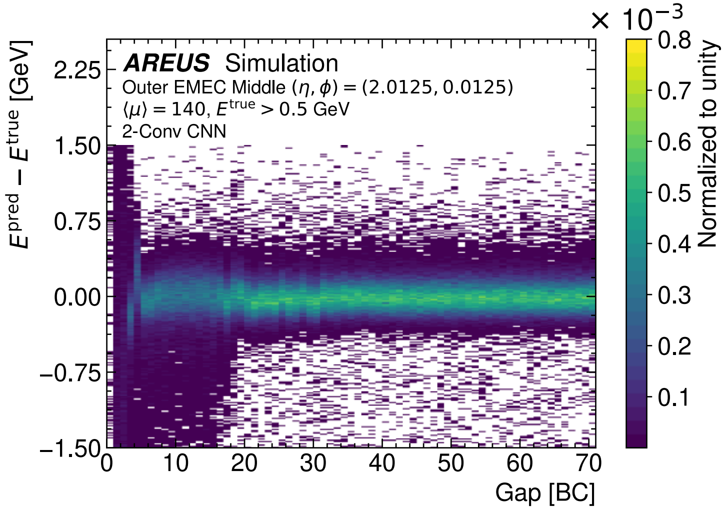
Performance Studies: Different Detector Regions





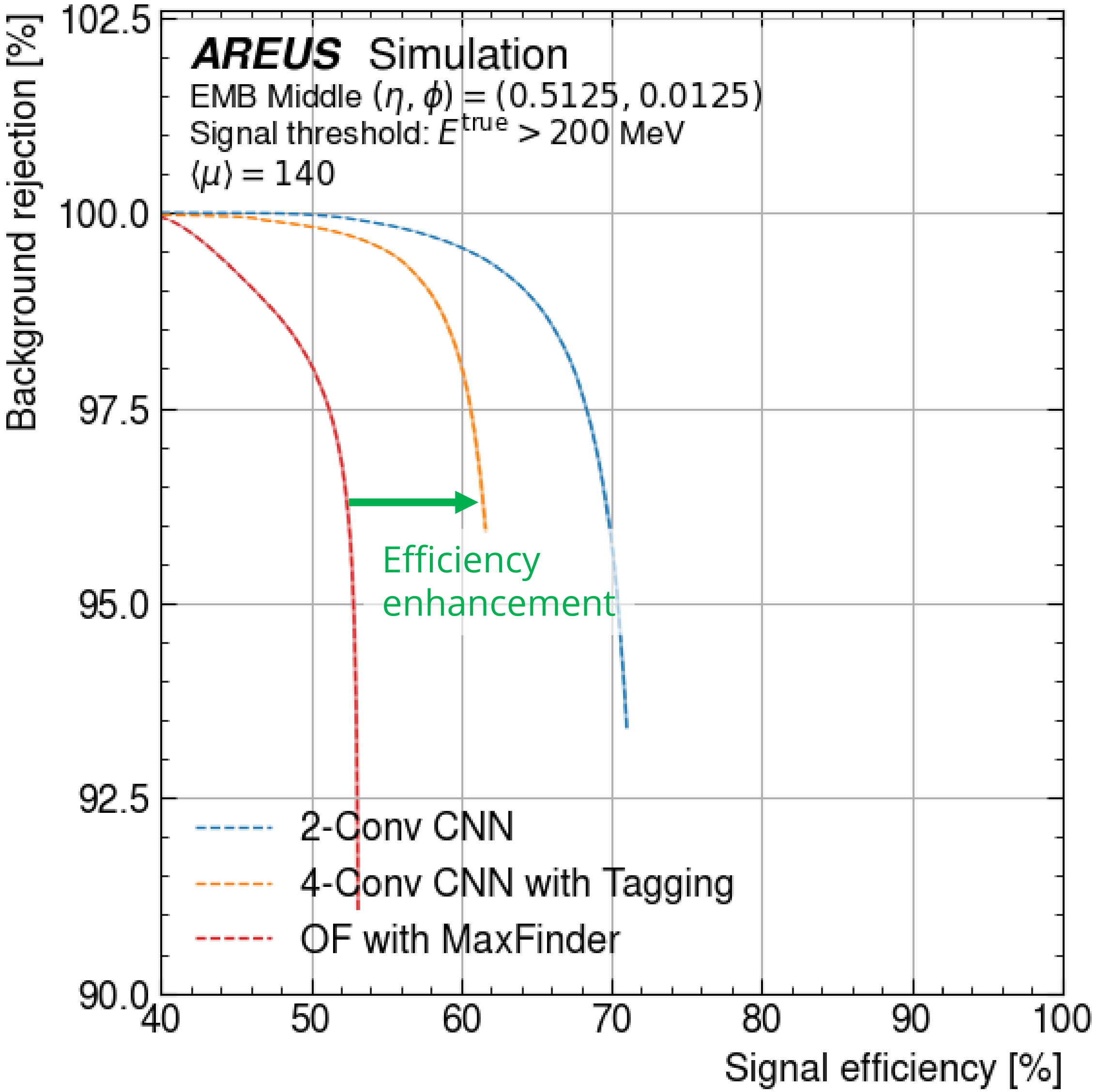






➢ Same architecture trained for different detector regions → shows similar results

Performance Evaluation: Energy Resolution



Receiver Operating Characteristic (ROC) Curves

- Indicate detection performance
- Signal efficiency $= \frac{true\ positives}{true\ positives + false\ negatives}$
- Background rejection $= \frac{true\ negatives}{true\ negatives + false\ positives}$
- Dependent on threshold

CNNs reach **higher signal efficiencies** at same background rejection level compared to OFMax

Performance Studies: Fakes

Spectrum of predicted transverse energy in BCs without energy deposition