Machine Learning-Based energy reconstruction for the ATLAS Tile Calorimeter at HL-LHC

Francesco Curcio on behalf of the Tile Calorimeter System

2nd EuCAIFCon, 16-20/06/2025

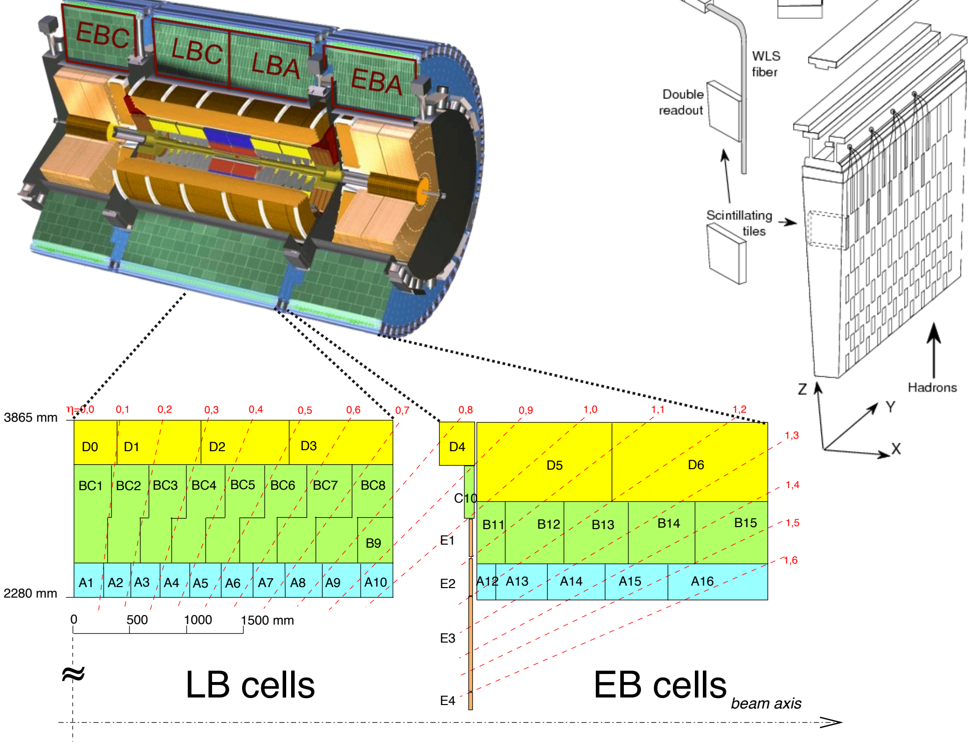
The ASFAE/2022/008 research projects acknowledge the financial support from the MCIU with funding from the European Union NextGenerationEU and Generalitat Valenciana.

Introduction

Tile Calorimeter

- Central hadronic calorimeter of the ATLAS experiment
  - Central Long Barrel (LBA, LBC) and Extended Barrel (two readout partitions, EBA and EBC)
  - 64 modules per partition
    - Up to 45 PMTs per module
- Sampling calorimeter composed of scintillators (active) and steel (absorber)
  - Charged particles produce light in plastic scintillators
  - The light is delivered to PMTs through WLS fibres
  - Reconstructs hadronic jets
  - Contributes to reconstruct the missing transverse energy
  - Input to trigger and muon identification
- Readout fibres groups into pseudo-projective cells
  - Each cell is read out by 2 PMTs
  - 5182 cells, 9852 PMTs

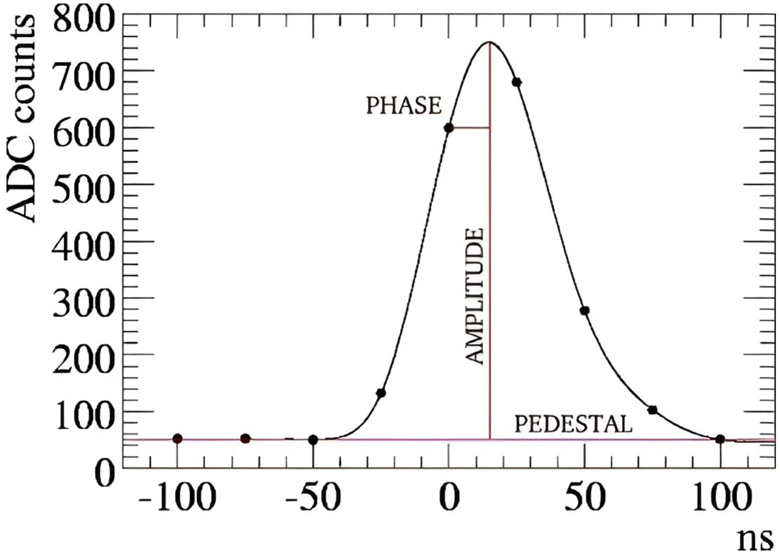
[Eur. Phys. J. C 84 (2024) 1313]



LBA/C	Long Barrel A/C
EBA/C	Extended Barrel A/C
PMT	Photo Multiplier Tube
WLS	Wave Length Shifting

Signal reconstruction

- In the TileCal readout, signals are sampled every 25 ns
- Each pulse is characterised by three main parameters:
  - Pedestal: the baseline ADC count value in absence of signal
  - Amplitude: proportional to the deposited energy
  - Phase: the time shift relative to the nominal bunch crossing
- In the legacy system, signals are processed online and offline using a 7-coefficient linear Optimal Filtering algorithm (OF)
  - Simple and fast response filter
  - Linear combination of the samples that uses weights determined from the known pulse shape and noise correlation matrix
  - Does not perform well under severe signal pileup conditions due to non-gaussian (asymmetric) components
- Reconstruction algorithms like OF or ML-based, aim to reconstruct the amplitude and the phase



Optimal Filtering

$$A = \sum_{i=1}^{n} a_i (S_i - p)$$

$$\tau = \frac{1}{A} \sum_{i=1}^{n} b_i (S_i - p)$$

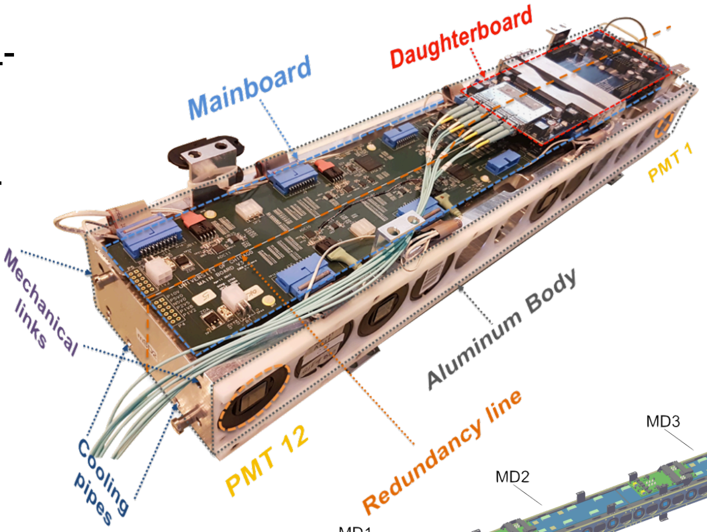
high and low gain (HG, LG) Amplitudes are measured in 2 gains, with a factor of 40 difference, both with 12-bit ADCs to maintain 10 MeV precision at energies < 10 GeV while avoiding saturation at high energies.

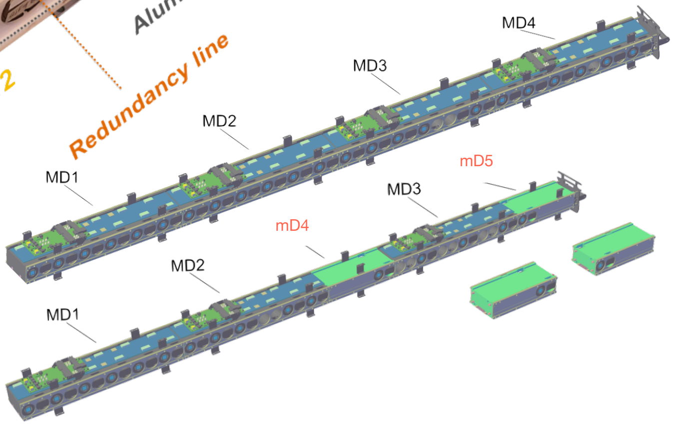
ADC Analog-to-Digital Converter

Tile Calorimeter - HL-LHC upgrade

- The upgrade of the ATLAS detector for the HL-LHC era aims to handle higher luminosity and data rates at HL-LHC
- Improved radiation hardness of electronics for high luminosity environment
- New mechanical frames to house on-detector electronics
- HV for each PMT will be created in service room ~100 m from the detector
- LV power supplies inside detector produce unified output voltage of 10 V
- Upgraded calibration systems
- Redundancy introduced in data links
- Replacement of 10% of the PMTs associated with the most exposed cells
- New readout, Front-end and Back-end electronics
  - Stream data from on-detector to new readout electronics at 40 MHz

HV	High Voltage
LV	Low Voltage
PMT	Photo Multiplier Tube
MD	Mini Drawer
mD	micro Drawer
SD	Super Drawer



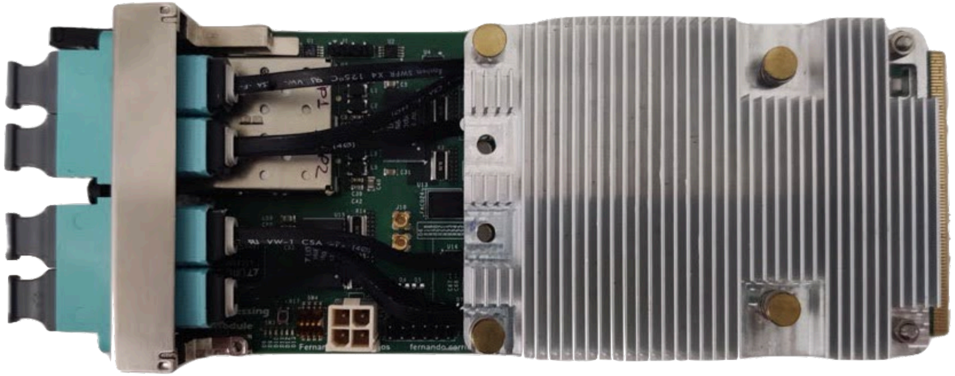


128 SDs for LB modules consist of 4 MDs (45 PMTs)

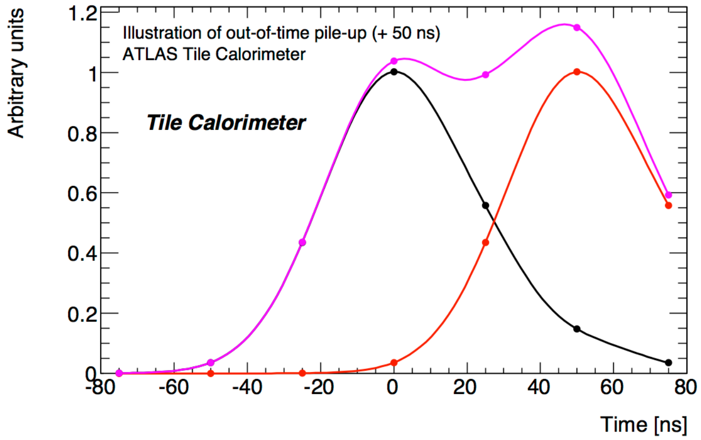
128 SDs for EB modules consist of 3 MDs + 2mDs (32 PMTs)

Motivation and goals

- During Phase-II, signals will be processed per bunch crossing (BC) and passed to the first level of trigger
  - Reconstruction will be done by FPGAs
- Higher pileup and rates, together with the need for real-time reconstruction, call for much more precise algorithms



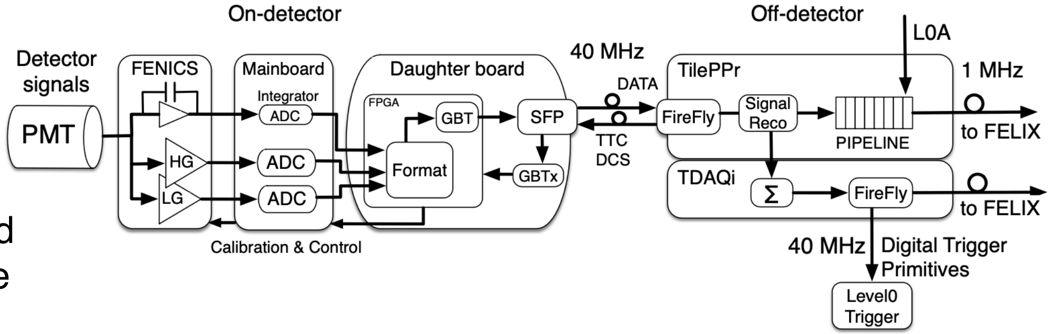
Compact Processing Module (CPM)

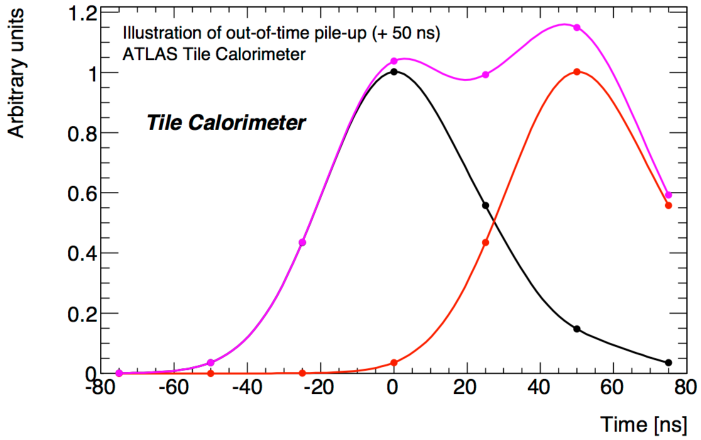


- Check the expected performance of different ML algorithms for HL-LHC conditions
  - Using PyTorch for this
- Optimise the number of parameters in order to be able to run the algorithm on FPGAs with the smallest possible latency

Motivation and goals

- During Phase-II, signals will be processed per bunch crossing (BC) and passed to the first level of trigger
    - Reconstruction will be done by FPGAs
- Higher pileup and rates, together with the need for real-time reconstruction, call for much more precise algorithms



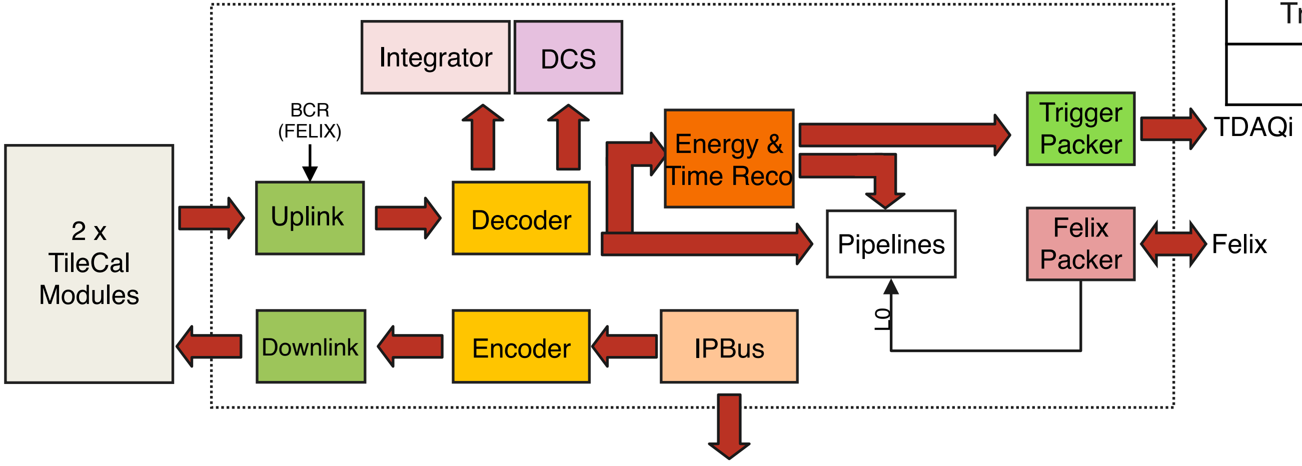


- Check the expected performance of different ML algorithms for HL-LHC conditions
    - Using PyTorch for this
- Optimise the number of parameters in order to be able to run the algorithm on FPGAs with the smallest possible latency

Signal reconstruction

- Xilinx Kintex UltraScale KU115
- Full firmware has 2x77 = 154 channels → 154 NN might be needed if 2 gains are used
- Clock domain crossing and synchronisation might be an issue
  - Passing data from 40/80/280 MHz

Firmware Block	Latency
Uplink	50 ns, 2 BC
Data Decoder	12.5 ns, 0.5 BC
Energy reconstrucion + sample delay	225 ns, 9 BC
Trigger Packer	12.5 ns, 0.5 BC
Trigger Interface	25 ns, 1 BC
Total	325 ns, 13 BC

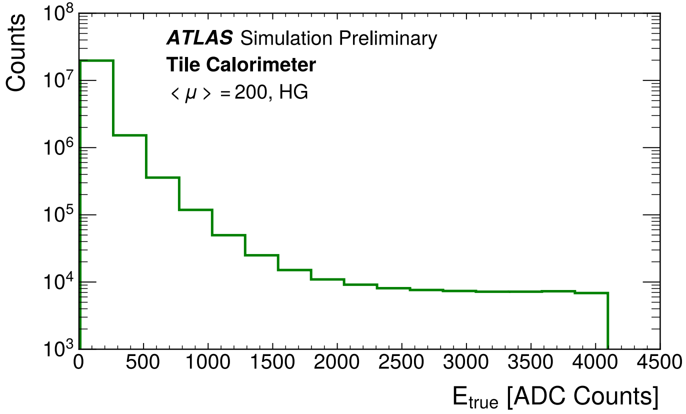


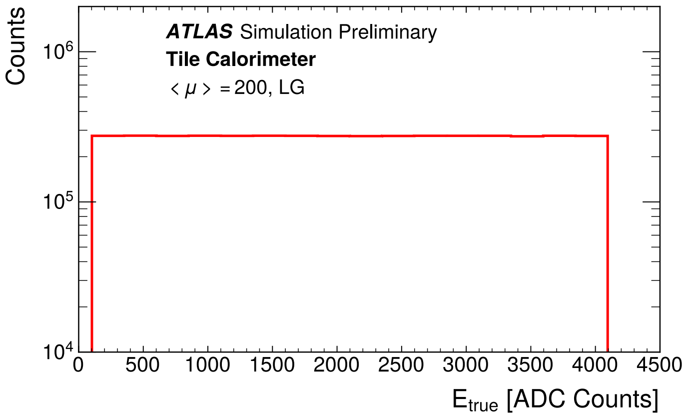
Block diagram of the Kintex Ultrascale firmware

BC	Bunch Crossing
NN	Neural Network
DCS	Detector Control System

Samples

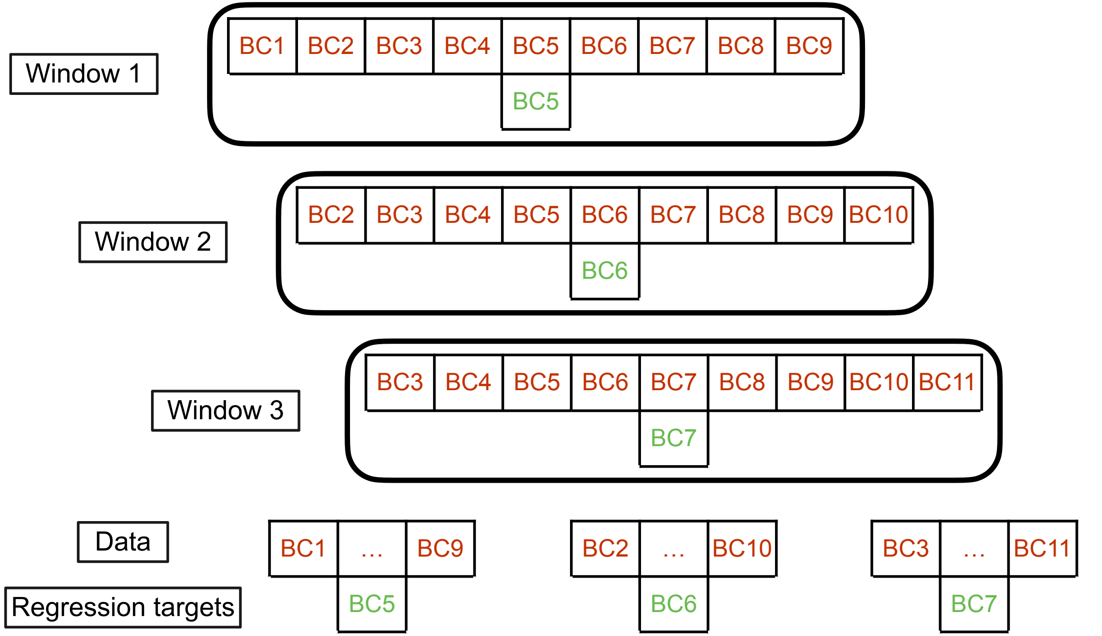
- The dataset is composed of ~1M consecutive bunch crossings with minimum bias average number of bunch crossings $\langle \mu \rangle$ = 200 with a superimposed flat distribution with a probability of 5%
    - ~1M x 64 modules x 4 channels = ~26.5M
- Only simulation of A1 cells of the Tile Calorimeter used
    - Different cells might require different models
- Samples have been divided in train (75%), validation (12.5%) and test (12.5%) sets
- The energy in ADC count for every bunch crossing is read in both gains
    - Models are trained on a mix between HG and LG
- After preprocessing, the samples are normalised in the range [0, 1]
- E_{true} is the regression target





Preprocessing

samples	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9	…	BCN
targets	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9	…	BCN



- For samples, $\mathrm{BC}_i \to E_{\mathrm{reco,i}}$
  - Simulated reconstructed energy
  - Inputs to our models
  - Sliding window of size 9
- For target, $\mathrm{BC}_i \to E_{\mathrm{true,i}}$
  - True energy for the i-th bunch crossing
  - Regression targets
  - Central energy of our window

BC Bunch Crossing

Preprocessing

samples	HG	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9
	LG	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9
	targets	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9

- If BC$_i$ saturates HG, take LG·40, else HG

samples	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9
targets	BC1	BC2	BC3	BC4	BC5	BC6	BC7	BC8	BC9

- E_{reco} is the simulated readout energy from the electronics.
- If any of the BC in the window saturates LG ($E^{\text{LG}}_{\text{reco}} > 4095$), drop the window.
- If any of the BC in the window has $E_{\text{reco}} \leq 10$ ADC counts, drop the window.
- If the central BC has $E^{\text{HG}}_{\text{true}} \leq 10$ or $E^{\text{LG}}_{\text{true}} \geq 4095$, the window is also dropped.

- For samples, HG, BC$_i$ $\rightarrow E^{\text{HG}}_{\text{reco,i}}$
  - Simulated reconstructed energy in the high gain reading
- For samples, LG, BC$_i$ $\rightarrow E^{\text{LG}}_{\text{reco,i}}$
  - Simulated reconstructed energy in the low gain reading
- For target, BC$_i$ $\rightarrow E_{\text{true,i}}$
  - True energy for the i-th bunch crossing
  - Values range from 0 to 40*4095

BC Bunch Crossing

Models

Models

Tested models

- Multi-layer Perceptron (MLP)
- 1D Convolutional Neural Network (CNN)
- Long Short-Term Memory (LSTM)
- Bi-LSTM
- $N_{\text{params}} \in [150{,}4000]$
- $N_{\text{inputs}} \in [15{,}13{,}11{,}9]$

Some of these dropped because

- Difficult to train
- Not easy to implement on FPGAs
- Suboptimal results
- Number of parameters too high
- High latency

After some optimisation converged on

- MLPs and CNNs
- $N_{\text{params}} \sim 150$
- $N_{\text{inputs}} = 9$

Machine Learning

- Supervised learning
- Feed-forward networks
- No dropout

Model architecture and loss

MLP

```
Sequential(
  (0): Linear(in_features=9, out_features=9, bias=True)
  (1): PReLU(num_parameters=6)
  (2): Linear(in_features=9, out_features=4, bias=True)
  (3): PReLU(num_parameters=4)
  (4): Linear(in_features=4, out_features=1, bias=True)
)
```

Total parameters: 148

CNN

```
Sequential(
  (0): Conv1d(1, 6, kernel_size=(3,), stride=(1,), padding=(1,))
  (1): PReLU(num_parameters=6)
  (2): Conv1d(6, 4, kernel_size=(3,), stride=(1,), padding=(1,))
  (3): PReLU(num_parameters=4)
  (4): Flatten()
  (5): Linear(in_features=36, out_features=1, bias=True)
)
```

Total parameters: 147

- Hybrid loss $= \alpha \cdot \frac{1}{N} \sum_i \left| y_i - \hat{y}_i \right| + \beta \cdot \sqrt{\frac{1}{N} \sum_i \left(y_i - \hat{y}_i \right)^2}$
  - Root Mean Squared Error (RMSE) used to keep same units as Mean Absolute Error (MAE)
  - Using $\alpha = \beta = 0.5$

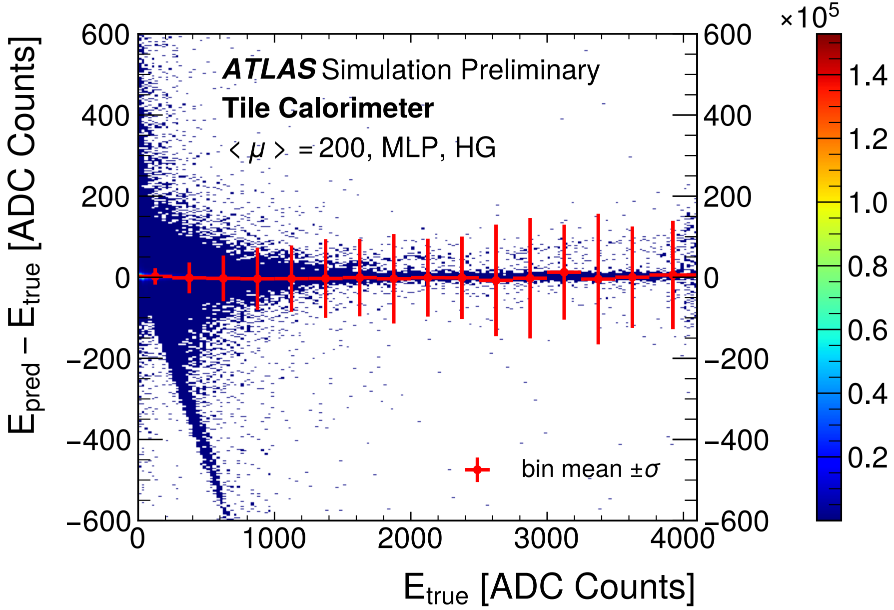
- Both architectures were determined trying not to create any bottlenecks and staying under $N_{\text{params}} \simeq 150$
- Undergoing further optimisation

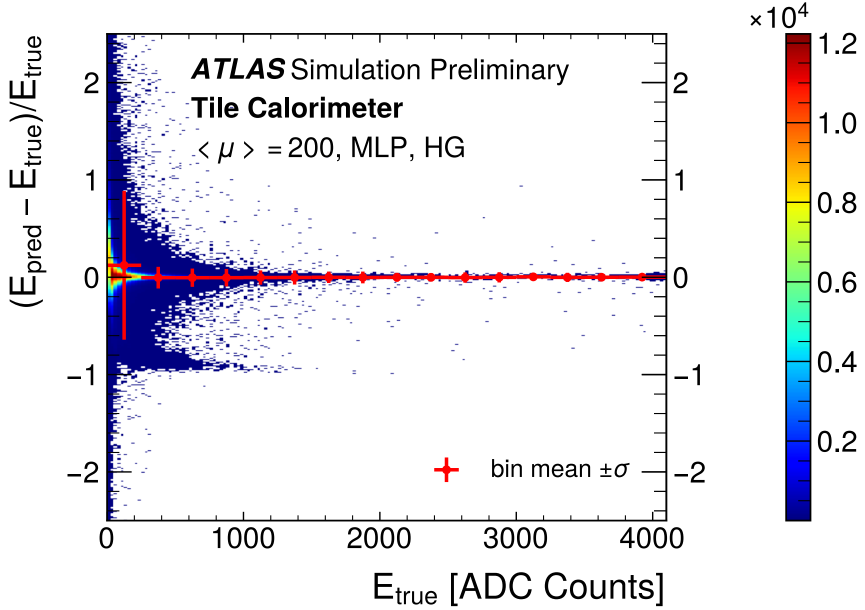
Results

Results - Plots

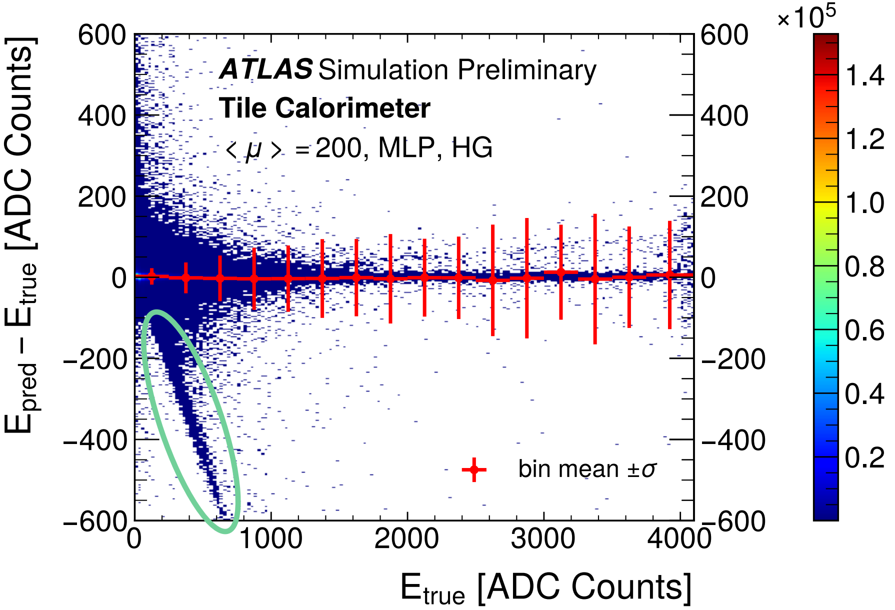
- Plots in the following slides are divided by model and gain
  - High Gain first for all models
  - Low gain in the second part
- Plots show 2D histograms as a function of the target energies E_{true} of:
  - Absolute error $E_{\text{pred}} - E_{\text{true}}$
  - Relative error $\dfrac{E_{\text{pred}} - E_{\text{true}}}{E_{\text{true}}}$
- Red markers show the average in absolute/relative error in each E_{true} bin with the correspondent standard deviation for the same bin

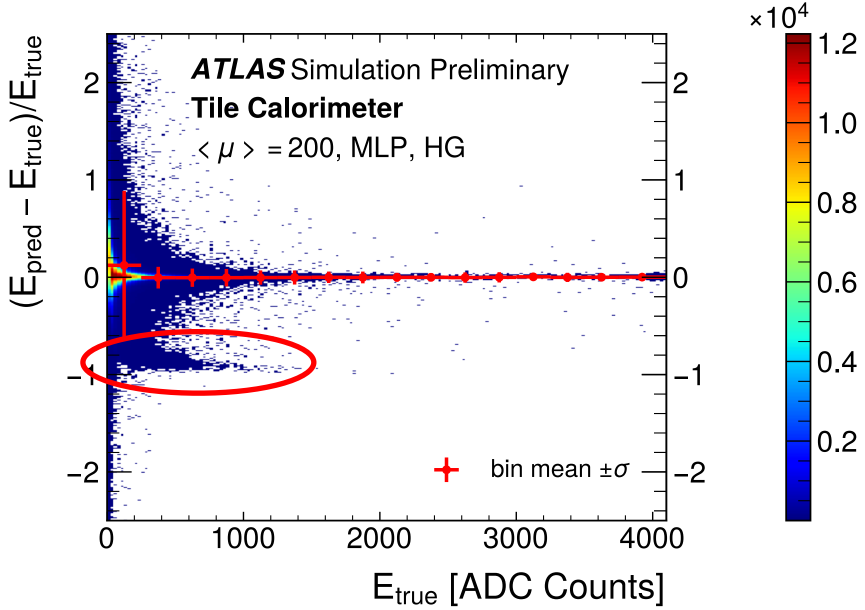
Results - MLP HG





Results - MLP HG



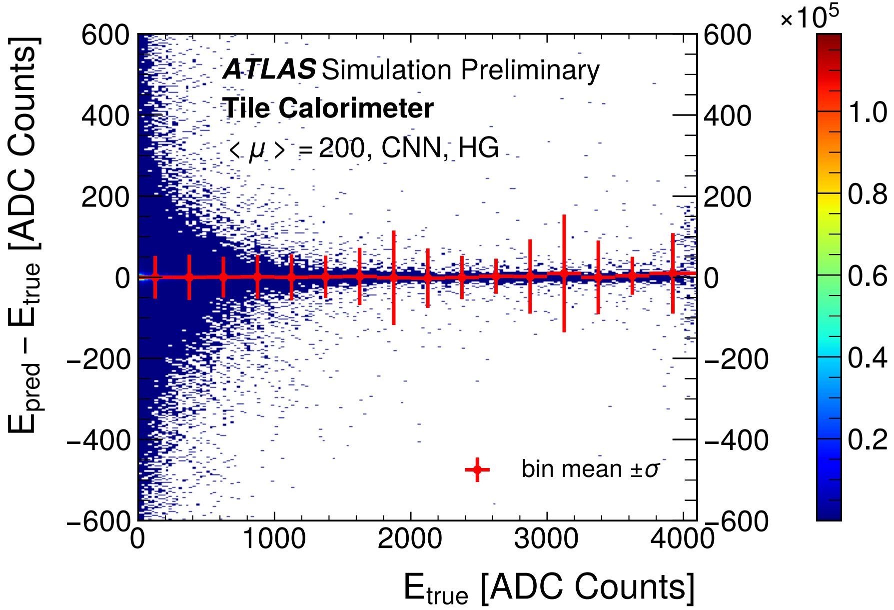


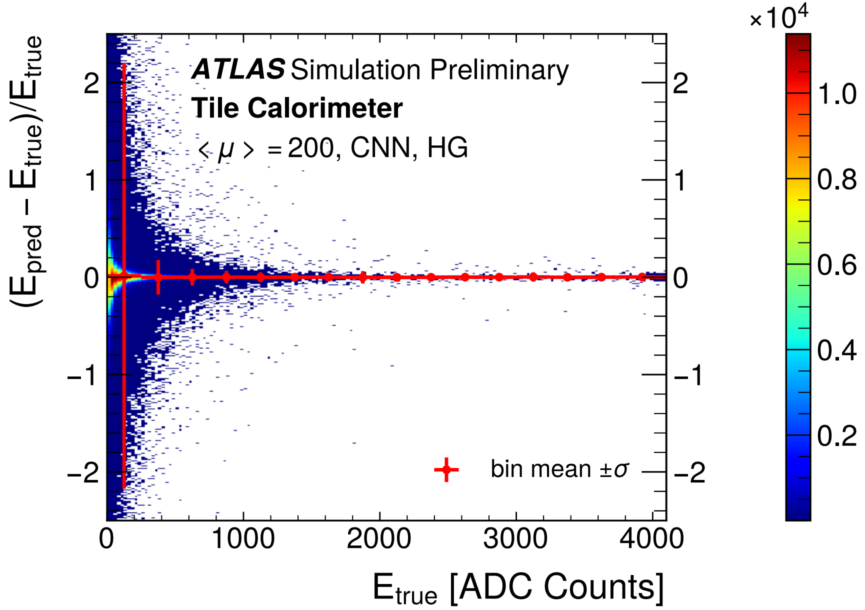
Diagonal line in bottom part of the plot is $E_{\text{pred}} \simeq 0$ or $E_{\text{pred}} - E_{\text{true}} \simeq -E_{\text{true}}$

$\sigma_{avg}^{err} = 99.76$ ADC Counts

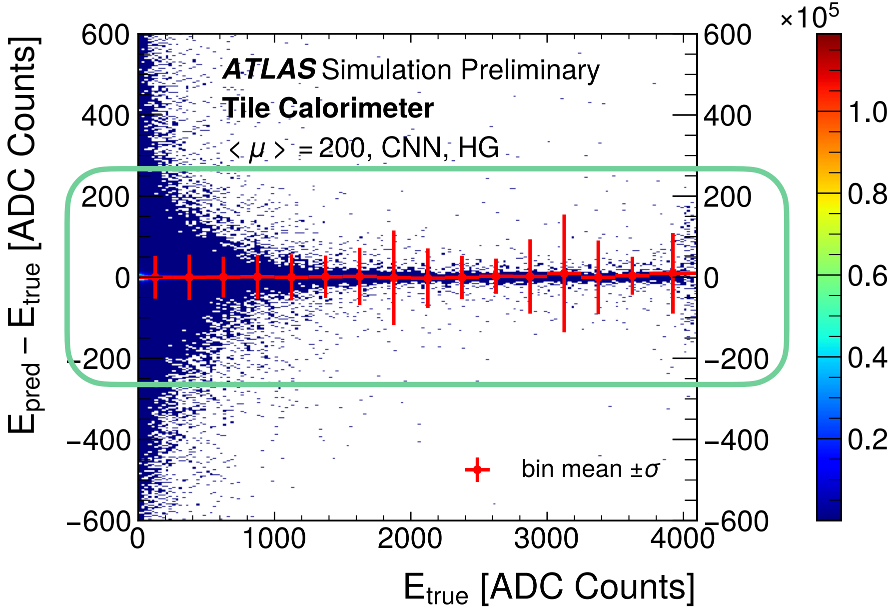
$$\frac{E_{\text{pred}} - E_{\text{true}}}{E_{\text{true}}} = -1 \rightarrow E_{\text{pred}} = 0$$

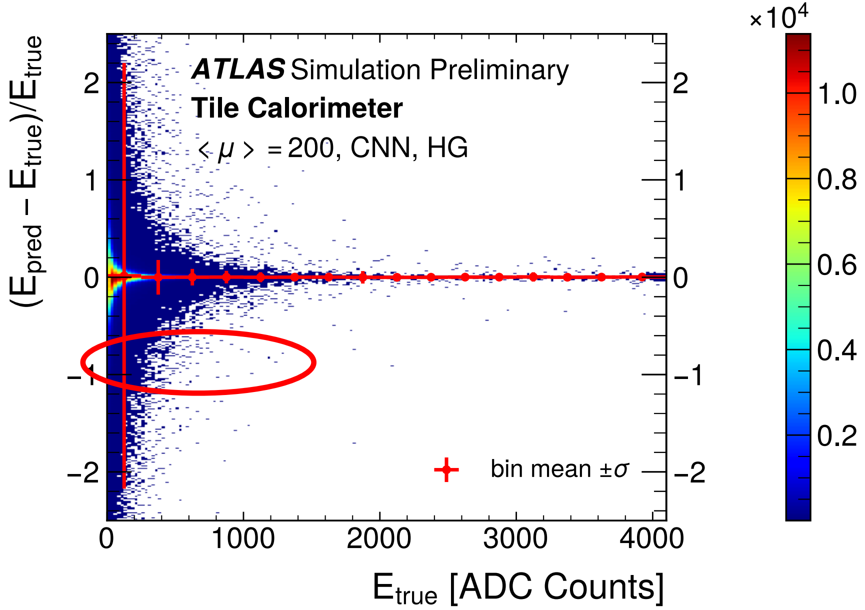
Results - CNN HG





Results - CNN HG



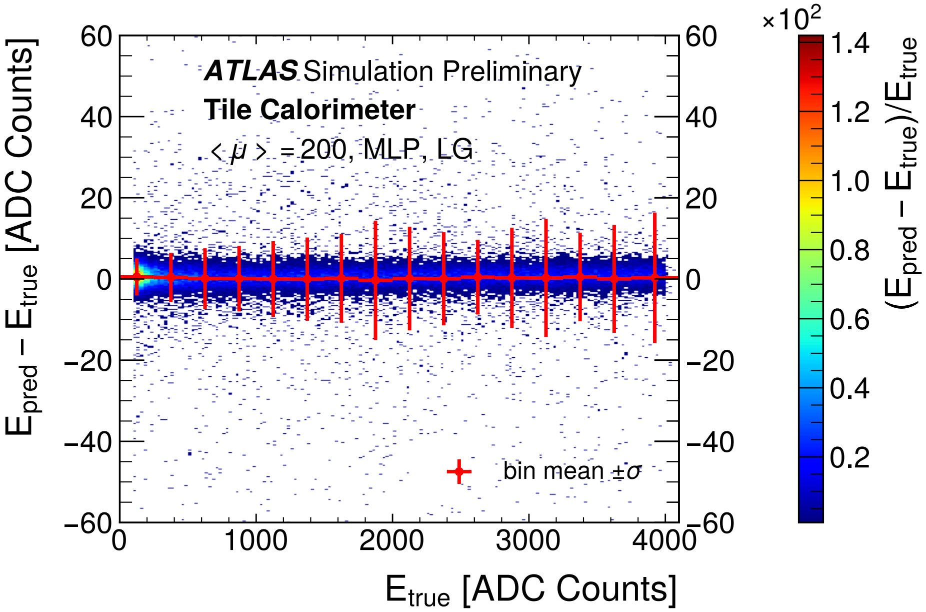


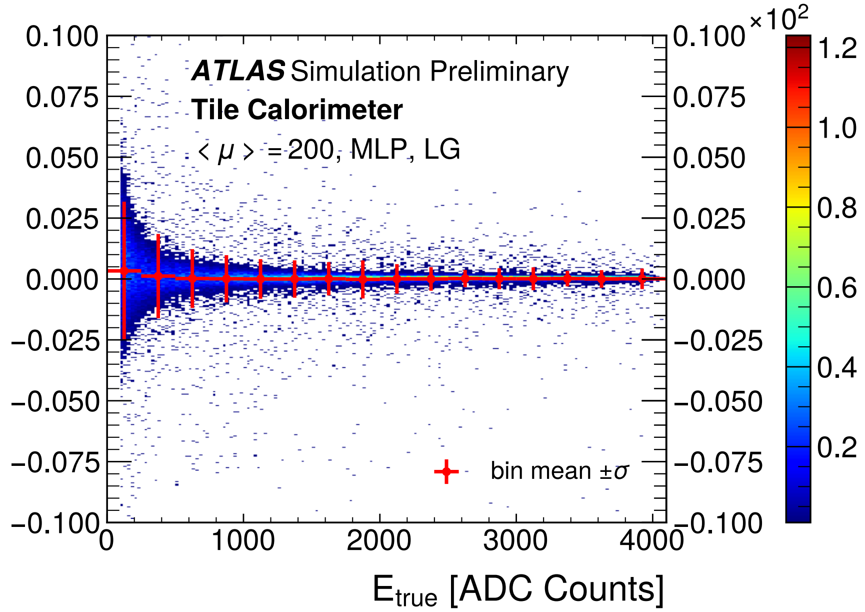
- Overall improvement in reconstruction
- CNN takes correlations between bunch crossings into account

$\sigma_{avg}^{err} = 72.12$ ADC Counts

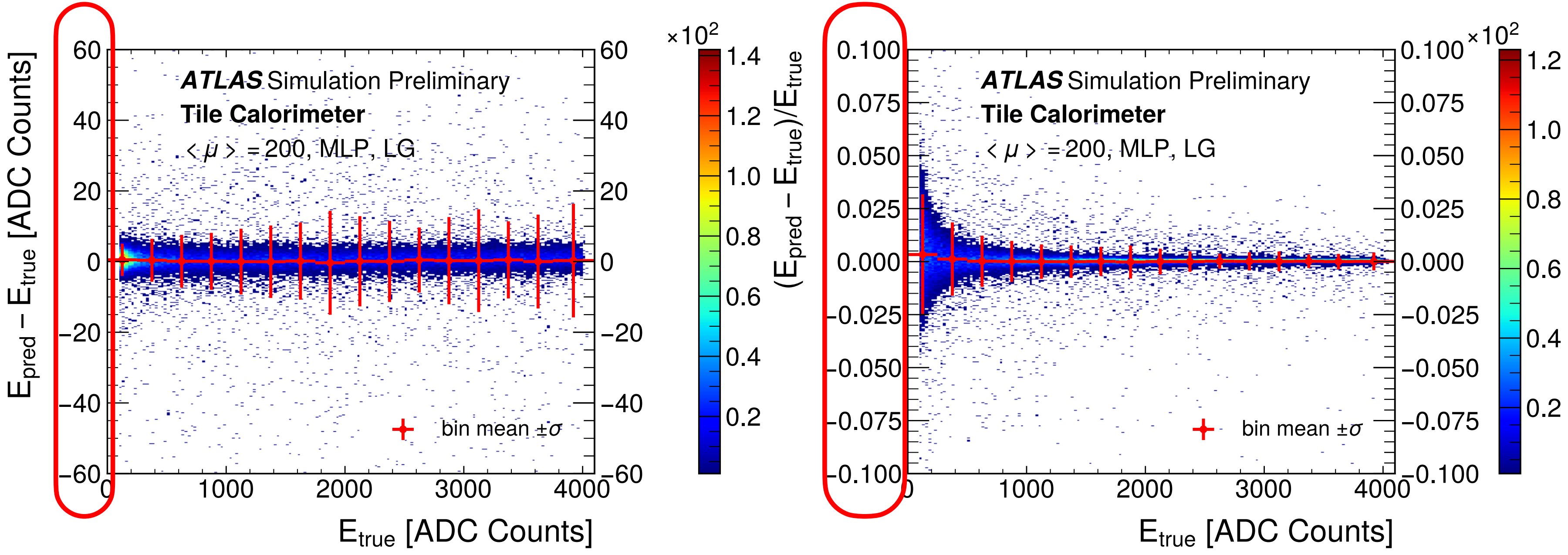
Step in $\frac{E_{\text{pred}} - E_{\text{true}}}{E_{\text{true}}} = -1$ gone

Results - MLP LG





Results - MLP LG

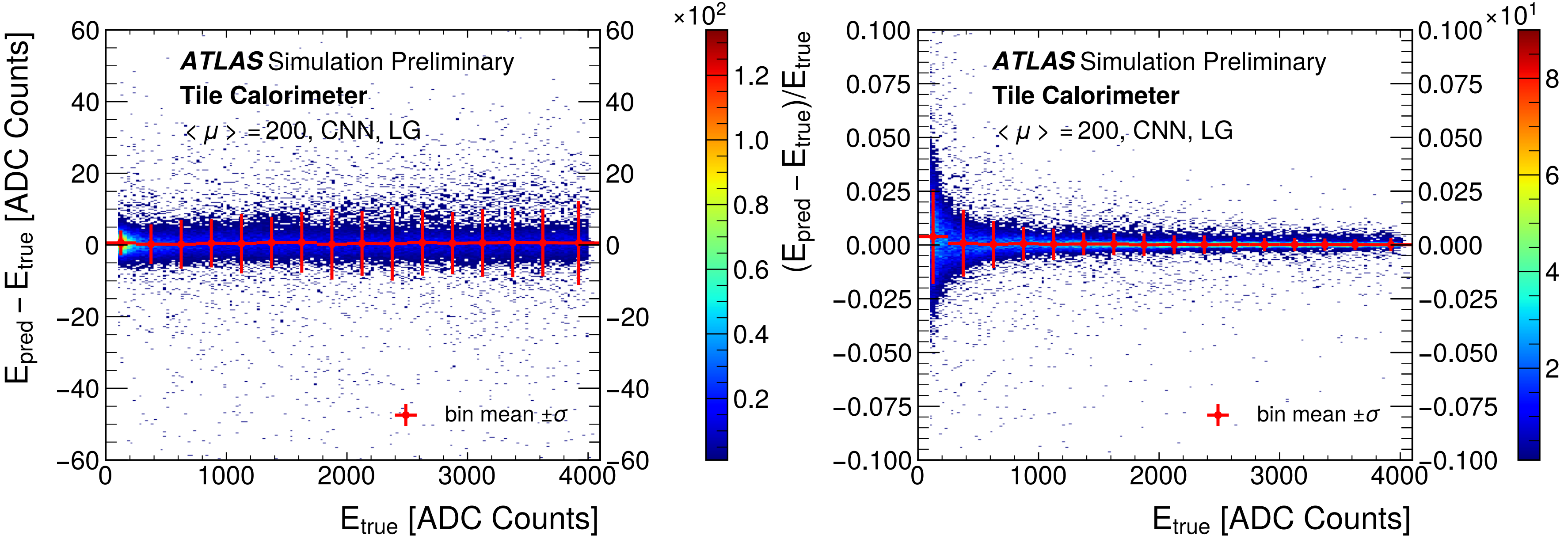


- Different scales with respect to HG
- Less noise and pileup less important in LG

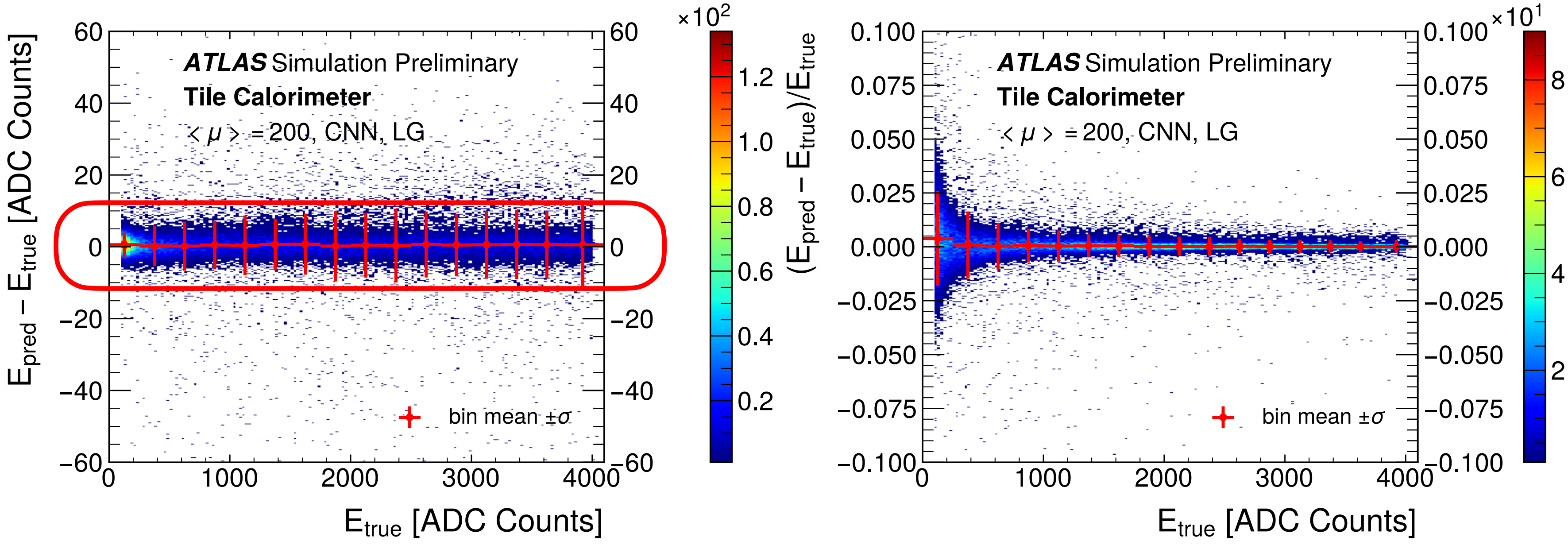
$\sigma_{avg}^{err} = 10.75$ ADC Counts

slight asymmetry in the bin-average distribution

Results - CNN LG



Results - CNN LG



better results than in the MLP case

$\sigma_{avg}^{err} = 8.36$ ADC Counts

slight asymmetry in the bin-average distribution

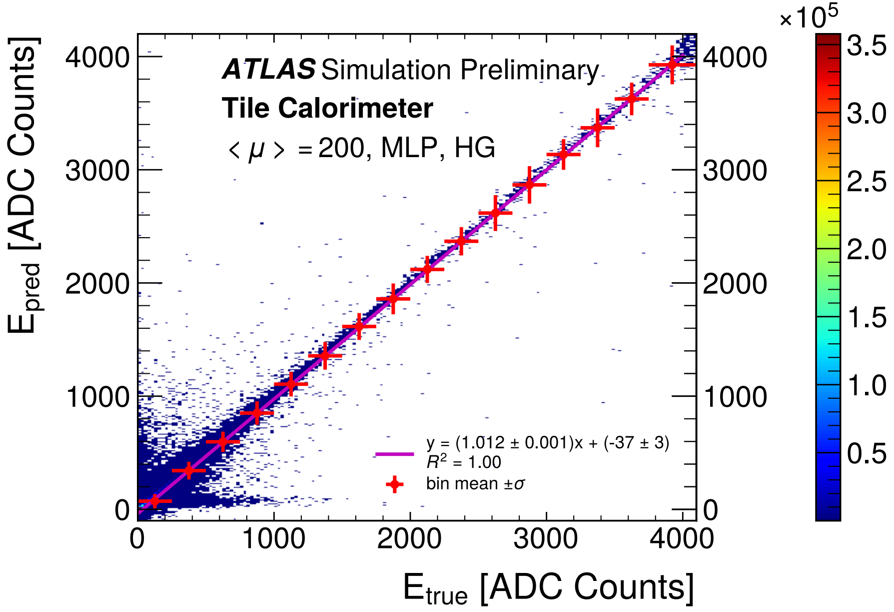
Summary and next steps

- HL-LHC conditions require real-time BC-wise energy reconstruction with high precision and low latency
- developing compact ML models suitable for FPGA deployment
- CNN outperforms MLPs in both gains
    - possibly because it takes into account correlations between different BC in the window
- CNN trained with hybrid loss ($0.5 \cdot$ MAE $+ 0.5 \cdot$ RMSE) gives best trade-off
- More checks on CNNs and different losses will be conducted in the future, stay tuned!

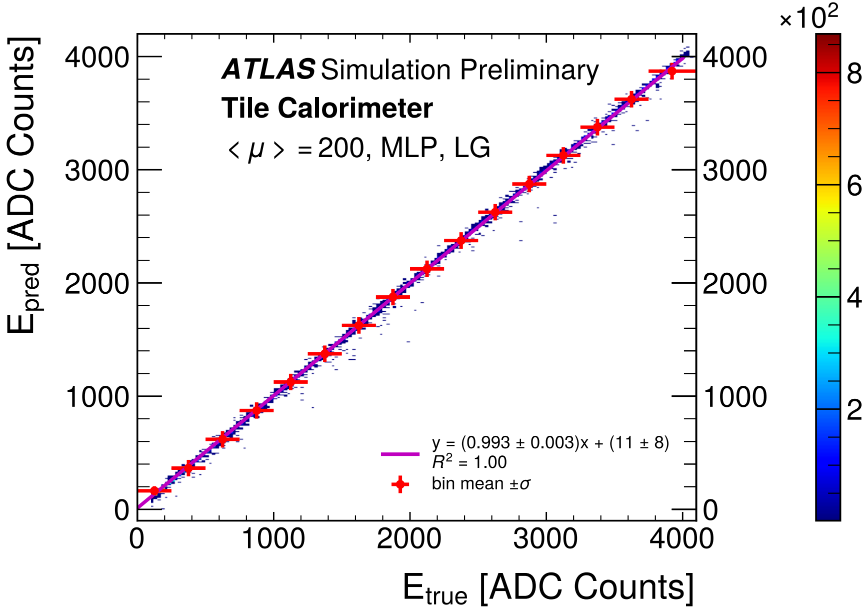
Thank you for your attention!

BACKUP

Results - MLP hybrid loss

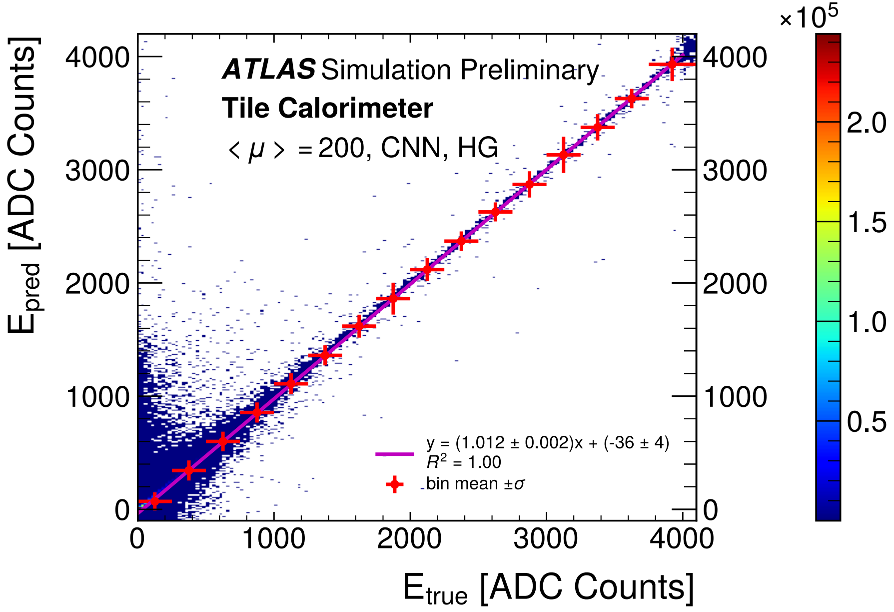


HG

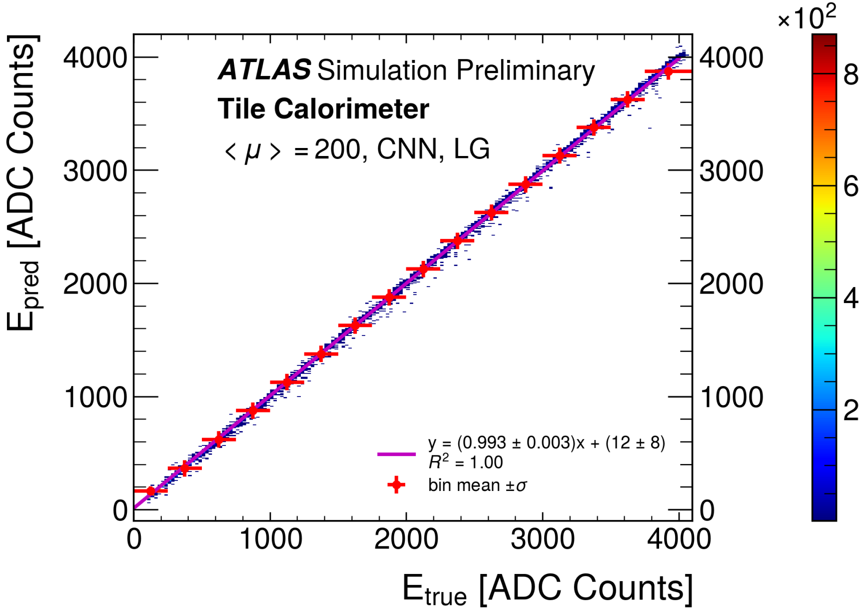


LG

Results - CNN hybrid loss

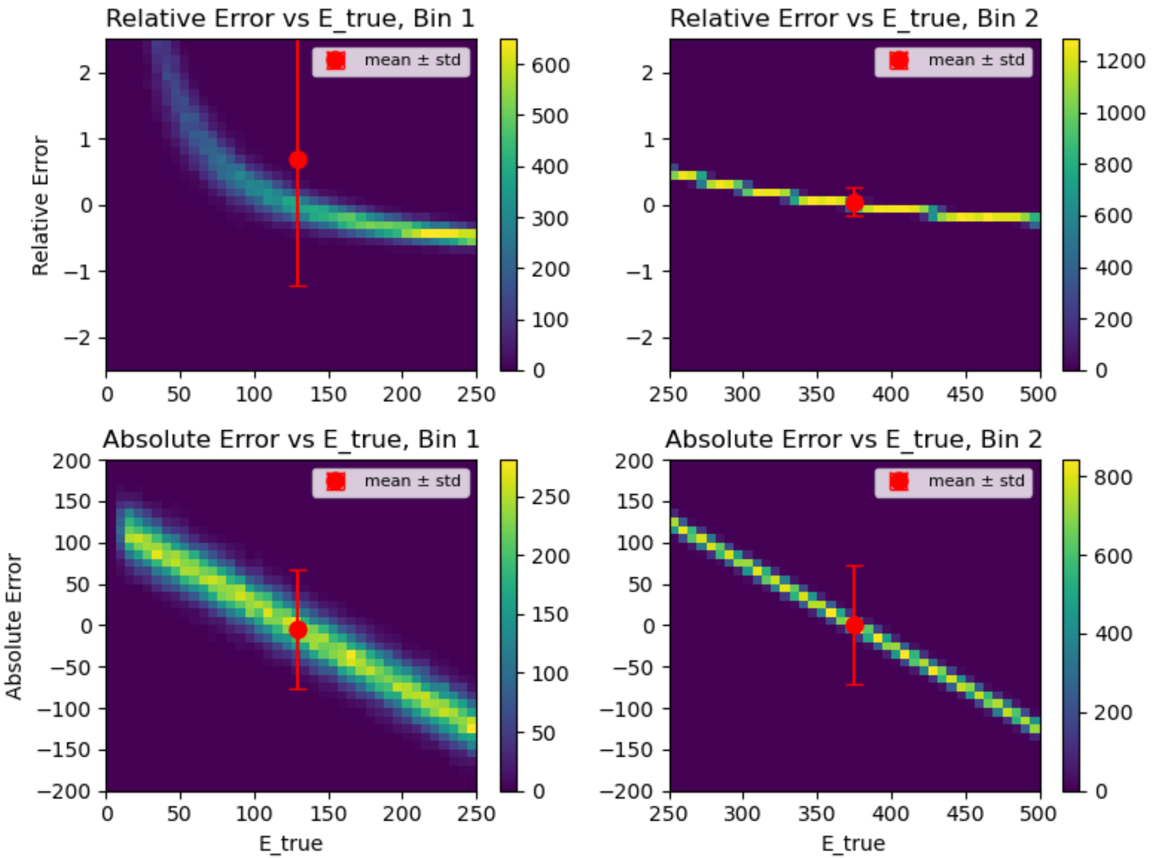


HG



LG

Different trend in relative and absolute errors



- Uniform E_{true} distribution between 0 and 500 ADC counts
- gaussian E_{true} distribution in each bin centred in the central E_{true} of that bin (125 and 325) with std = 50 and 20, respectively
- NB not real data, just toy model
- no inference from the model