Energy Reconstruction with Recurrent Neural Networks

IN2P3/IRFU Machine Learning workshop

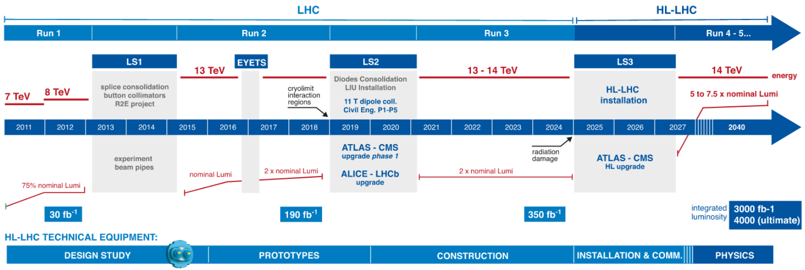
Georges Aad, Thomas Calvet, Nemer Chiedde, Etienne Fortin, **Lauri Laatu**, Emmanuel Monnier

16.03.2021

Content

The Phase-II Upgrade of the LHC

Upgrade of the ATLAS experiment

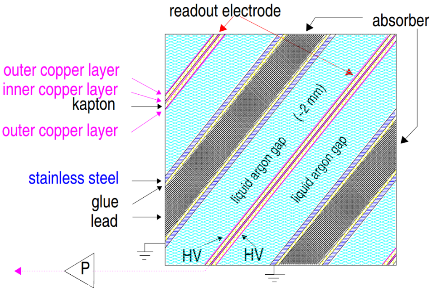


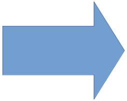
- The High Luminosity LHC (HL-LHC) is a important milestone for particle physics
  - Increase the luminosity to study rare processes
  - Increase the collision rate to up to 200 simultaneous p-p collisions (pileup) per bunch crossing (BC)
- The detectors will be upgraded to cope with the high collision rate at the HL-LHC
  - In particular the ATLAS calorimeter readout electronics will be completely replaced

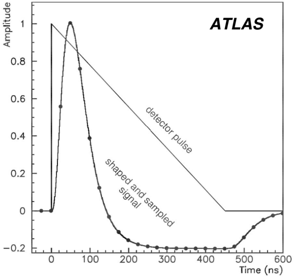
Liquid Argon Calorimeter

Energy reconstruction in the LAr calorimeter

- Liquid Argon Calorimeter (LAr) mainly measures the energy deposited by electromagnetically interacting particles
  - Consisting of ≈ 180 000 calorimeter cells
- Passing particles ionize the material
  - Bipolar pulse shape with total length of up to 750 ns (30 BCs)
  - Pulse is sampled and digitized at 40MHz
- Energy reconstruction is done real-time and used in triggering decision
  - Using the digitized samples from the pulse







Energy Reconstruction

Energy reconstruction in the LAr calorimeter

- Current energy reconstruction uses optimal filtering algorithm with maximum finder (OFMax)
  - Using five samples around pulse shape peak
  - Assuming perfect pulse shape
- High pileup leads to higher rate of overlapping pulse shapes
  - Distorted bipolar shape → significantly decreased performance of OFMax

- Energy is computed real-time at 40MHz
  - Need to use electronic boards based on FPGAs
- Phase-II electronics with high-end FPGAs
  - Increased computing capacity
  - Improved online energy reconstruction using machine learning based methods
- Constraints from running on FPGAs
  - Latency, frequency and occupancy
  - See next presentation by Etienne Fortin

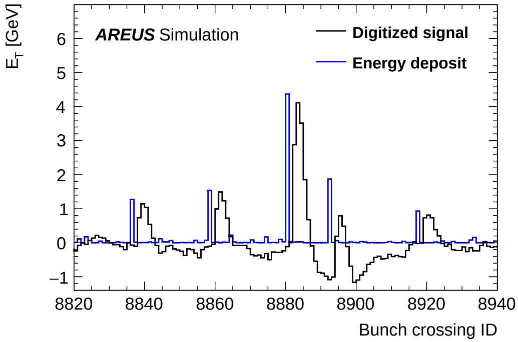
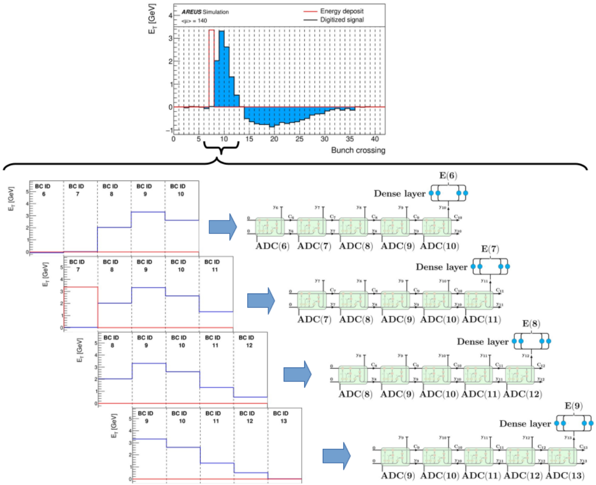


Table of Contents

LSTM Network

Using a many-to-one network for energy reconstruction

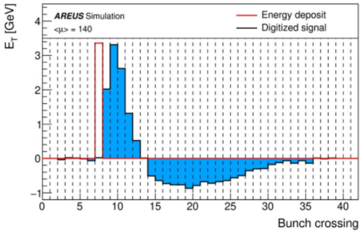
- Recurrent Neural Networks (RNNs) are designed to process time series data
    - Long Short-Term Memory (LSTM) network for efficiently handling past information
    - LSTM consists of neural network layers that process the new time input combined with past processed state
- Use digitized samples as inputs for the recurrent network
- Full sequence split into overlapping subsequences with a sliding window
- One energy prediction per subsequence



LSTM Network Without State Resets

Using a many-to-many network for continuous energy reconstruction

- Use the LSTM cell to process all digitized samples in one continuous chain instead of a sliding window
- Apply the same LSTM operation for each bunchcrossing combining the past state and new ADC value
- Use each intermediate state for energy reconstruction
- No state resets (stateful in Keras)
- Reduces the computational requirements when each prediction requires only one iteration of the cell



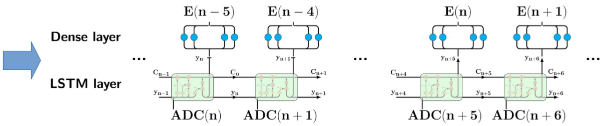
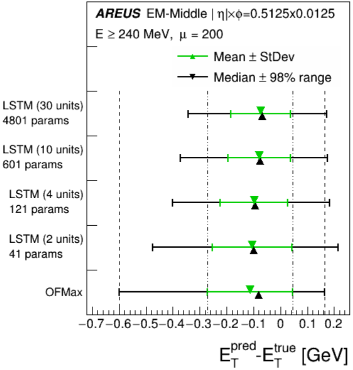


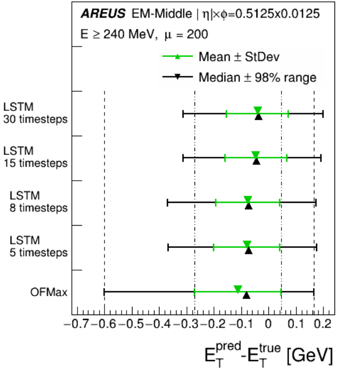
Table of Contents

LSTM Network Optimization

Find the smallest well performing network, example for sliding window LSTM

- Use standard deviation and 98% range to compare energy resolution
  - Non-gaussian distribution of the energy resolution
- Optimization of the energy resolution while keeping the network size under control
  - Vary the network parameters: internal dimension (units), sliding window size (timesteps)
  - Network trained with simulated data of a single LAr calorimeter cell using the AREUS software

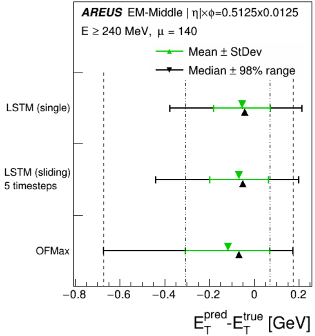


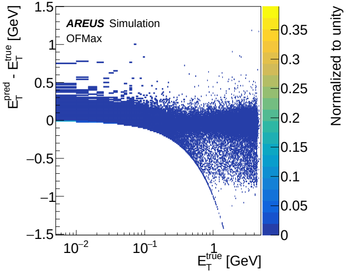


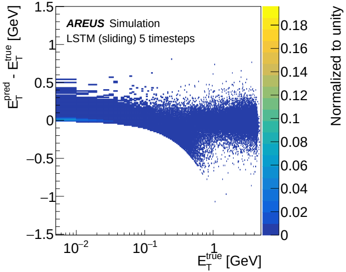
LSTM Performance

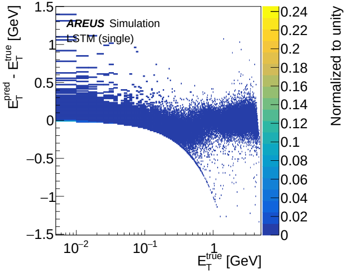
Energy resolution in comparison to OFMax

- Both LSTM architectures perform better than OFMax
- LSTM with sliding window is more resilient to outliers and focuses on the peak
- Stateful LSTM performs slightly better overall with better correction of long range effects but has a wavy feature









LSTM Performance

Resolution as a function of gap to previous energy deposit in BCs

- Clear performance decrease with OFMax at low gap
- LSTM with sliding window showing slight underestimation of energy at low gap and overestimates at around a gap of 20 BC
  - Increasing the amount of timesteps removes both effects
- Single cell LSTM handles long range effects better

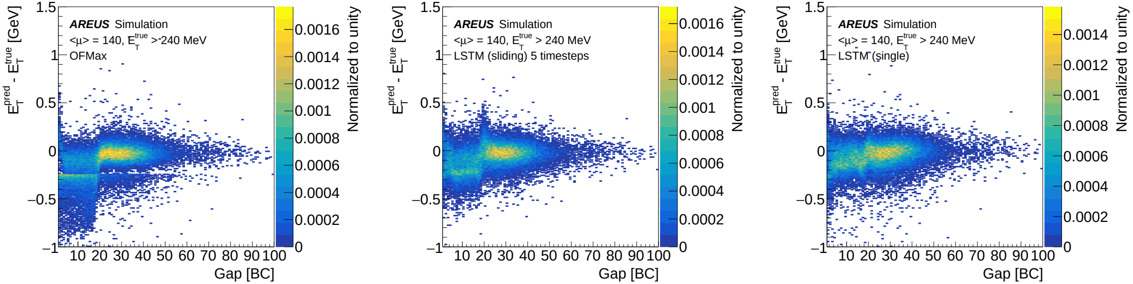


Table of Contents

Conclusion

Energy reconstruction using recurrent neural networks

- Energy reconstruction with LSTM overperforms legacy algorithms in Phase-II conditions
  - Better energy resolution overall
  - Better recovery of energy resolution with overlapping signals
- Strict resource and latency constraints limit the size of the networks
  - Energy resolution optimized while keeping the network size small
- FPGA implementation of LSTM presented next by Etienne Fortin
- Next steps:
  - Using other RNN architectures (Vanilla RNN and GRU)
  - Quantization aware training using QKeras (better adapted to FPGA implementation)
  - Continue studies of the robustness against HL-LHC conditions (bunch train structure, varying instantaneous luminosity, ...)