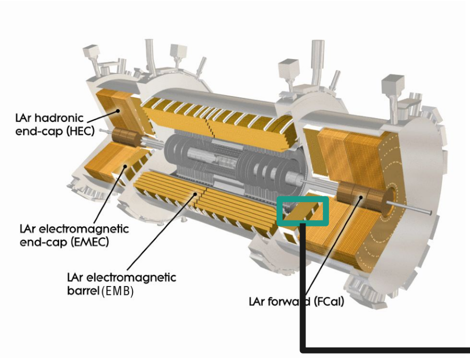
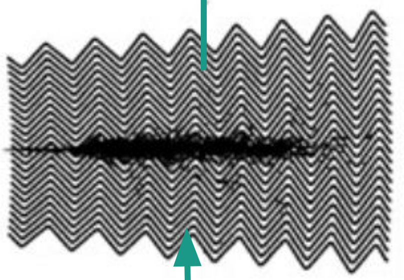
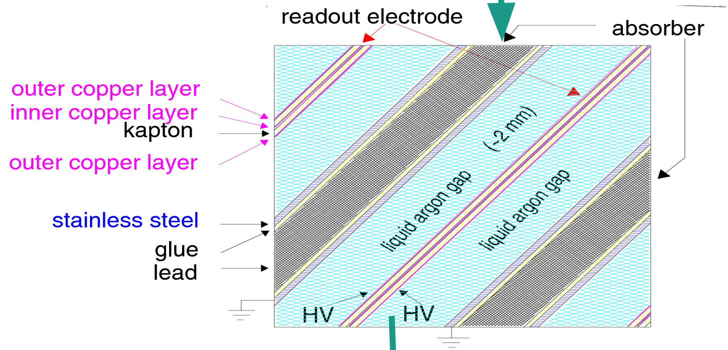
ATLAS LAr : Modèles et démonstration

Raphaël BERTRAND
Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

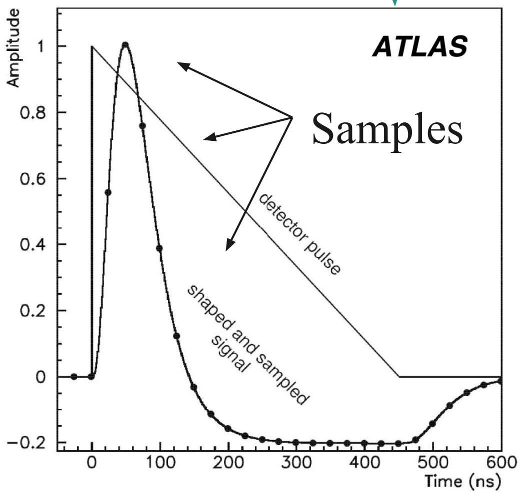
The Liquid Argon calorimeter (LAr)







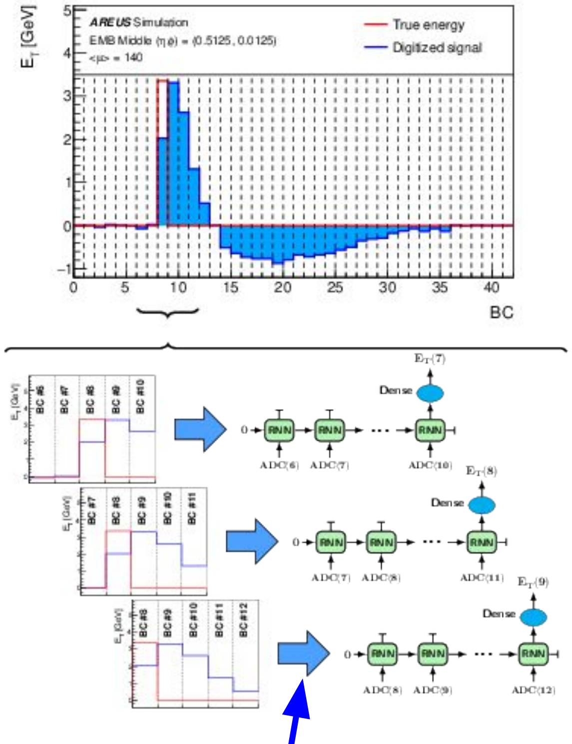
- Measure the energy of (mainly) electromagnetic interacting particles
  - Liquid argon as active medium
  - Lead/copper/tungsten as absorber
  - 180000 channels
- Electronic signal amplitude proportional to the deposited energy in the calorimeter
  - Shaped and sampled at 40 MHz
- Trigger capabilities



Comput Softw Big Sci **5**, 19 (2021)

RNN structure

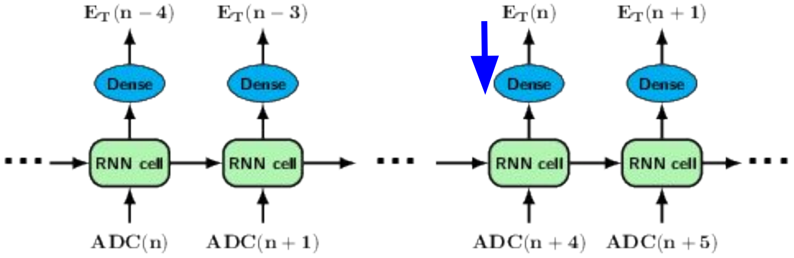
- Single cell architecture
  - Full history learned
  - Less robust against intermittent problems such as noise bursts
  - Need large cells to handle full history
- Sliding window architecture (retained)
  - Learn only local effects (what we need)
  - Intermittent problems have only short time effect
  - Suitable for small cells



<u>Single cell architecture</u>

Continuous computation with a single cell

Takes into account full past info (from the beginning of run)



<u>Sliding windows architecture</u>

Computation on a moving slice (fixed intervals)

Takes into account a limited set in the past

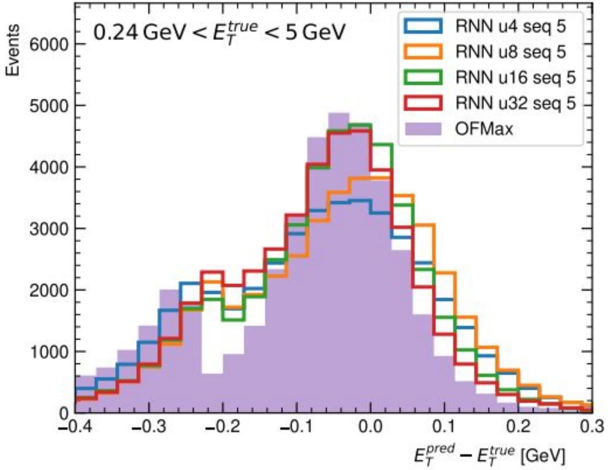
(1 sample in the past for this example)

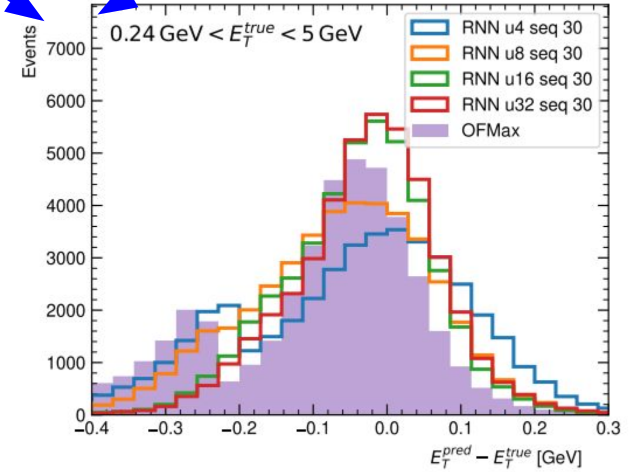
RNN Performance

- Compare energy resolution between RNNs and OFMax
  - RNNs with increased size
  - Keep size under control to fit FPGAs
- Second peak in resolution due to overlapping events
- Use Std. Dev. as metric (although the shape is not very gaussian)

Number of units (u)

Sequence length (seq)

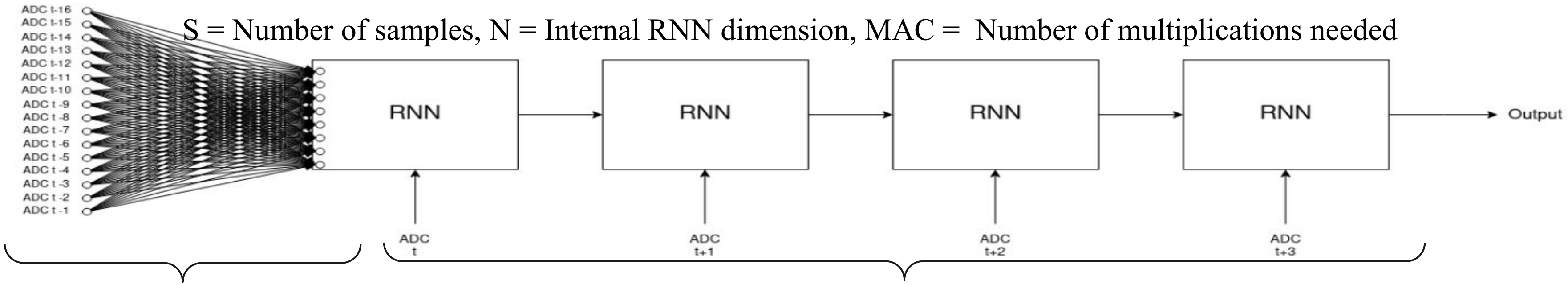




Lauri Laatu, PhD thesis (2023)

Optimisation of computational resources

S = Number of samples, N = Internal RNN dimension, MAC = Number of multiplications needed

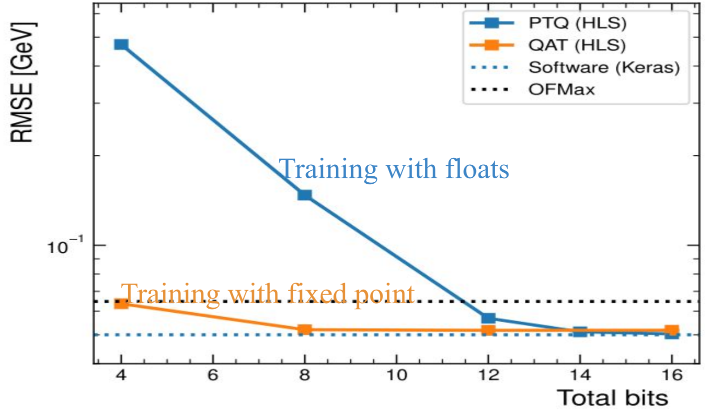


Dense layer : MAC $\propto$ S×N

RNN layers : MAC $\propto$ S×N^2

- Long sequences needed to efficiently correct for pileup
  - Significant computational resources needed for RNN cells
- Replace RNN cells in the past by a dense layer
  - Dense to correct to pileup, RNN to compute the amplitude
  - Reduce the number of needed multiplications by a factor 4
    - For a network with dimension 30 and sequence length 20
  - No effect on performance
- Reduce number of bits needed for arithmetic computation
  - Replace floating point with fixed point operation
  - Train the network directly with fixed point (QAT)
  - Quantization aware training (QAT) can reduce the number of needed bits by a factor 2

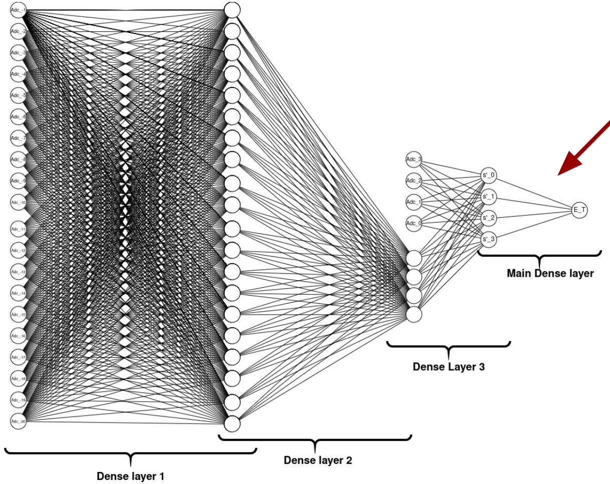
Simulation of the energie resolution in firmware as function of the number of bits



Two Stage Dense Architecture

- Dense architectures are the ones that need the least processing power
  - But their are concern about the latency
- Dividing the dense into two stages
  - First to compute the corrections from the past
    - Can be done before the energy deposit
  - Second compute the pulse amplitude (thus the energy) including the corrections from past events
- Preliminary results are encouraging
  - Almost the same resolution at RNN with less MAC units
  - Fresh results that still need confirmation

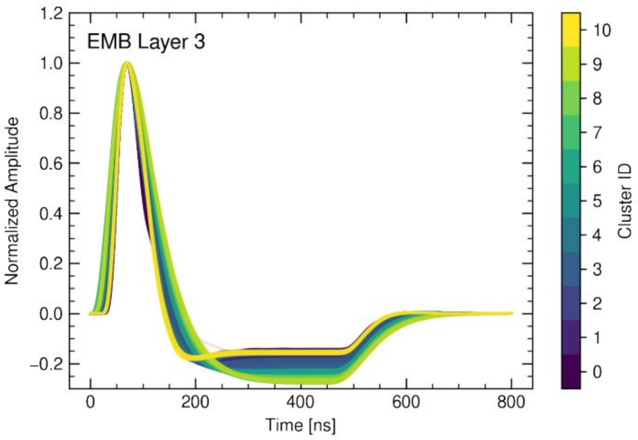
Dense with only samples
from the past
No impact on latency

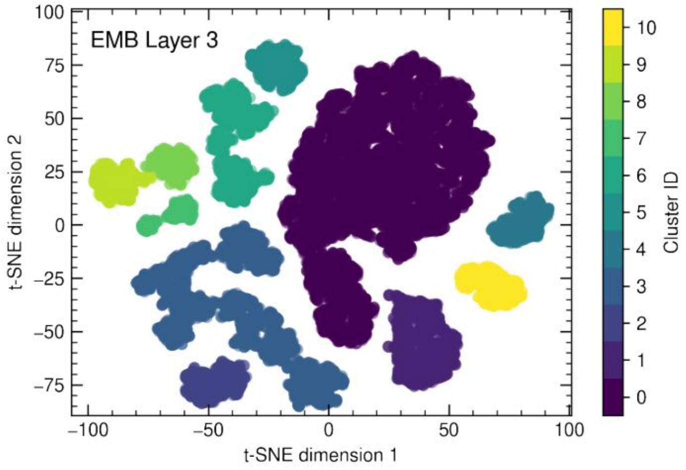


Dense with samples on the pulse
Mixed with corrections from the
first dense

From single cell to the full detector

- Training 180 000 NNs is not a raisonnable task
  - Not just CPU/GPU but also need to validate then
- Group cells with “similar” pulse shape into a single NN
- Cells are grouped using an unsupervised clustering method
  - t-SNE to reduce the dimensionality: from n samples on the pulse to 2 dimensions
  - DBSCAN to cluster in two dimensions





From single cell to the full detector

- Clusters manage to catch the geometric symmetries in the detector
  - Symmetry in Phi (Azimuthal angle)
  - Changing cell size (capacitance) and thus pulse shape in Eta (~ Polar angle)
- Confirmed clustering does not degrade RNN performance
  - Same resolution training on cells from the same clusters
  - Dramatic degradation of resolution if training on a random cell outside the cluster
    - Training on all clusters at once does not recover the performance