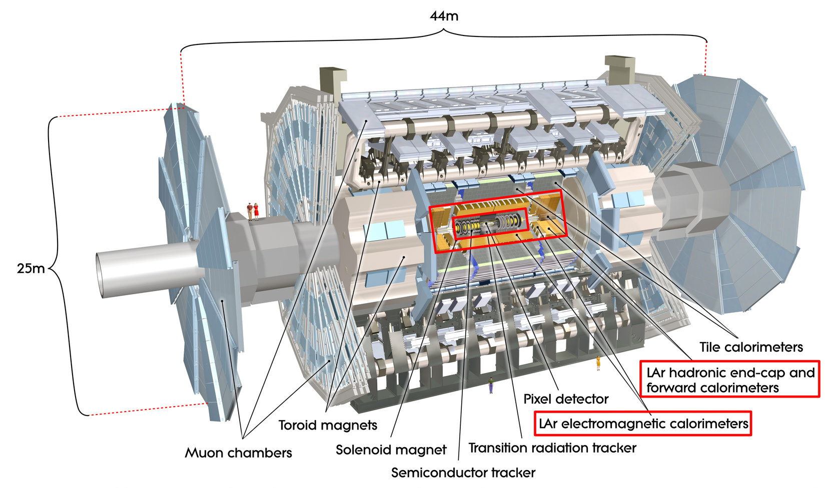
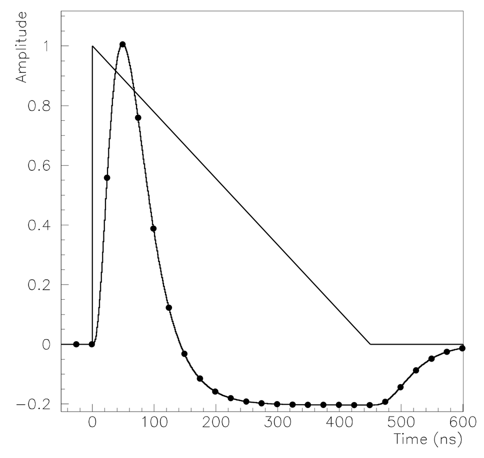
Machine Learning for Real-Time Processing of ATLAS Liquid Argon Calorimeter Signals with FPGAs

DIS2022: XXIX International Workshop on Deep-Inelastic Scattering and Related Subjects

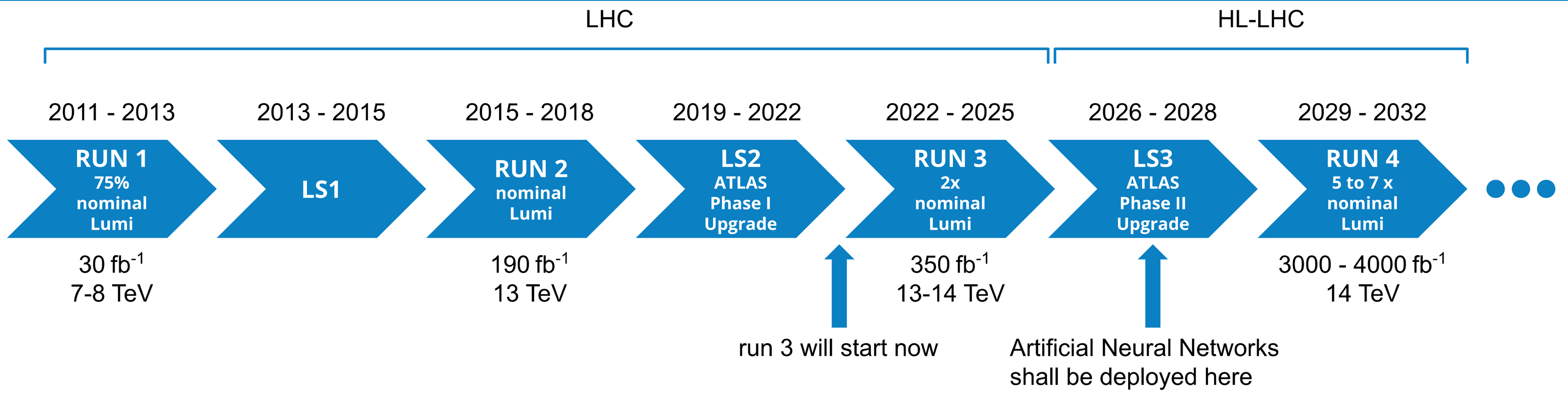
1. ATLAS Liquid Argon Calorimeter





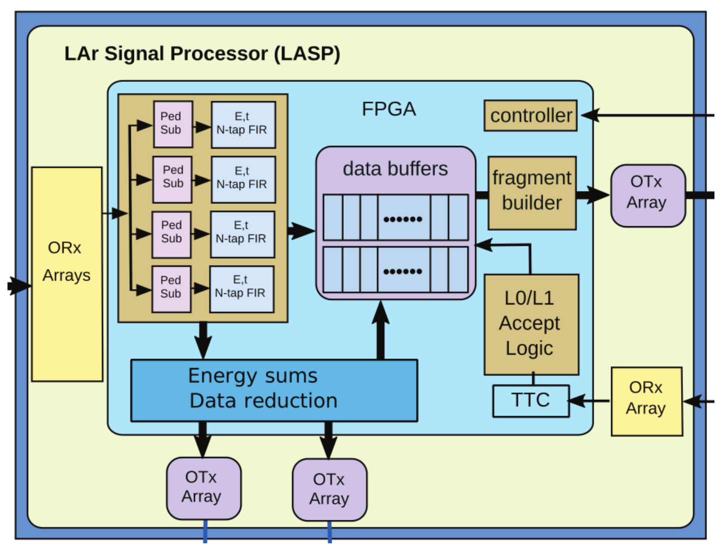
- Sampling calorimeters for the measurement of energy deposited by electrons, photons and hadronic jets
- 182 000 cells filled with Liquid Argon as medium for ionization
- Resulting triangular pulse is amplified, shaped and digitized at 40 MHz
- Energy reconstruction with **Optimal Filter (OF)**: linear combination of up to 5 samples

2. Phase II Readout Electronics Upgrade



- High Luminosity LHC (HL-LHC) planned to start in 2029 with up to 7.5 times nominal luminosity
- About **140 proton-proton-collisions per bunch crossing**
- Challenges for LAr calorimeter readout under HL-LHC conditions:
  - Revised trigger scheme allows selection of events at 1 MHz and in subsequent bunch crossings (BC)
  - Number of events with **overlapping signals will increase** → in-time and out-of-time pile-up

- New Liquid Argon Signal Processor (LASP) boards will be installed during Long Shutdown 3 (LS3, Phase-II upgrade):
- **FPGA to implement more complex real-time energy reconstruction algorithms**
- Maximum latency of ~150 ns required by trigger
- Up to 512 detector cells processed on one FPGA



- Readout chain can be simulated with AREUS software:
  - Supports electronics noise, digitization and LHC bunch train structure
  - Generates digitized pulse sequence as expected from specific detector cell

3.1 Convolutional Neural Networks (CNN)

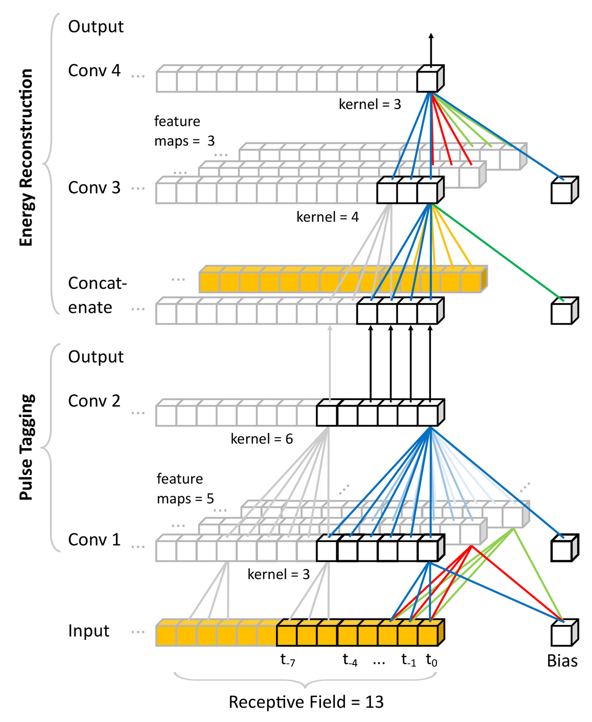
- **Machine Learning** solutions under investigation to **replace OF** in future HL-LHC conditions

- **Pulse tagging sub-network** (2 layers)
  - Sigmoid activation function

- **Energy reconstruction sub-network** (1-2 layers)
  - Uses results of tagging sub-network and raw ADC samples
  - ReLU activation function

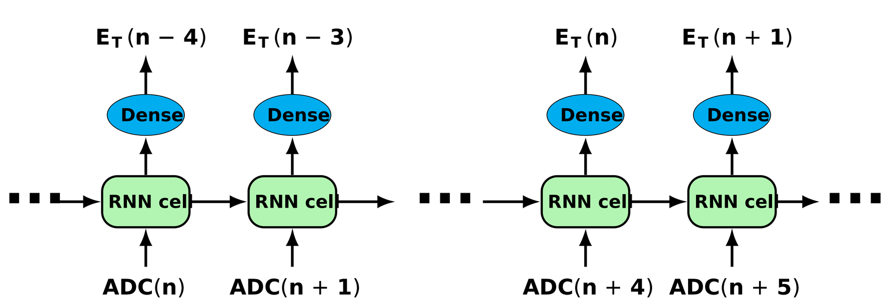
- Trained and evaluated on AREUS samples

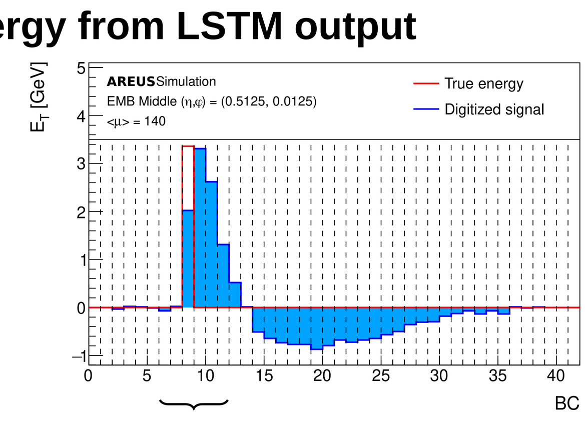
- Training in two stages:
  1. Tagging part only as pre-training
  2. All layers together for energy reconstruction



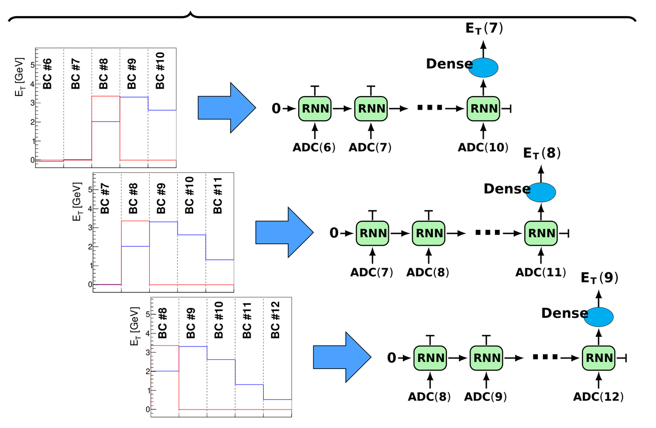
3.2 Recurrent Neural Networks (RNN)

- RNNs have connections between nodes over time → memory
- Long Short-Term Memory (LSTM) well suited for long sequences due to gated design
- Limited number of internal dimensions and only one layer due to FPGA resource constraints
- **Single dense neuron** as decoder to **reconstruct energy from LSTM output**





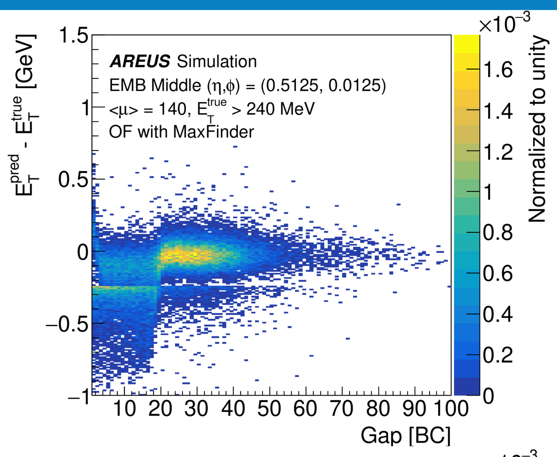
- Single-cell LSTM (above)
  - Operates sample per sample **on entire sequence**
  - Expected to be more robust for overlapping pulses
- Sliding-window LSTM (right)
  - **Newly instantiated** LSTM operates on fixed number of **past samples at each BC**
  - Expected to be more robust for isolated pulses
- Vanilla RNN: simple structure to reconstruct energy and forward information

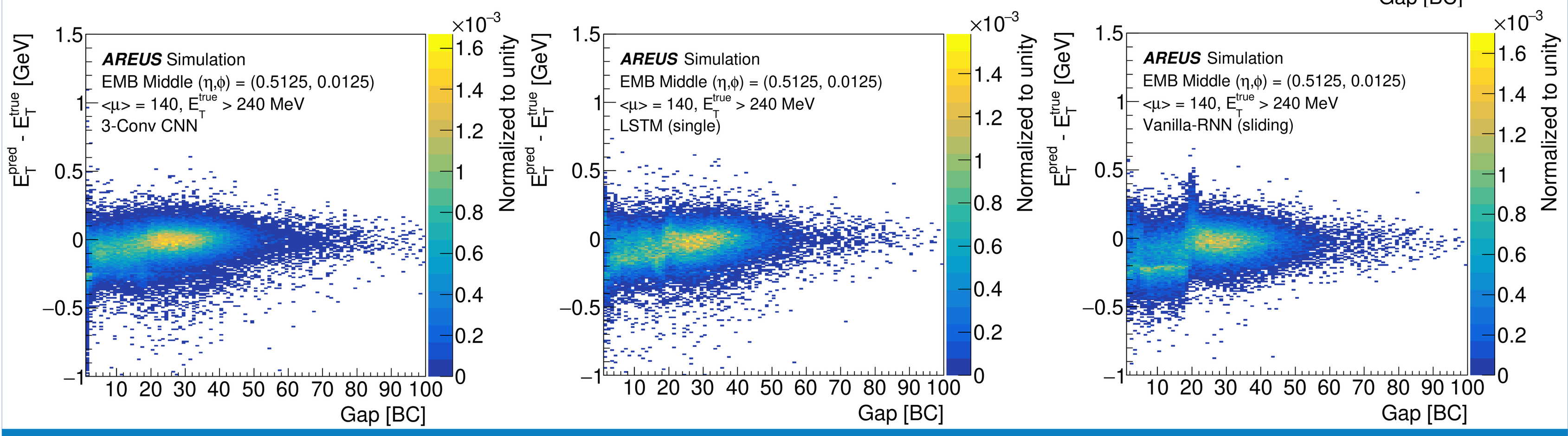


3.3 Performance in Simulation

- CNNs and RNNs **outperform OF** in both mean and bias behaviour
- CNNs/RNNs show **less biased** behaviour with **overlapping pulses**

Figures: difference between predicted and true energy over the temporal gap between two higher energy deposits for OF (right) and and different types of artificial neural networks (ANN) (bottom)



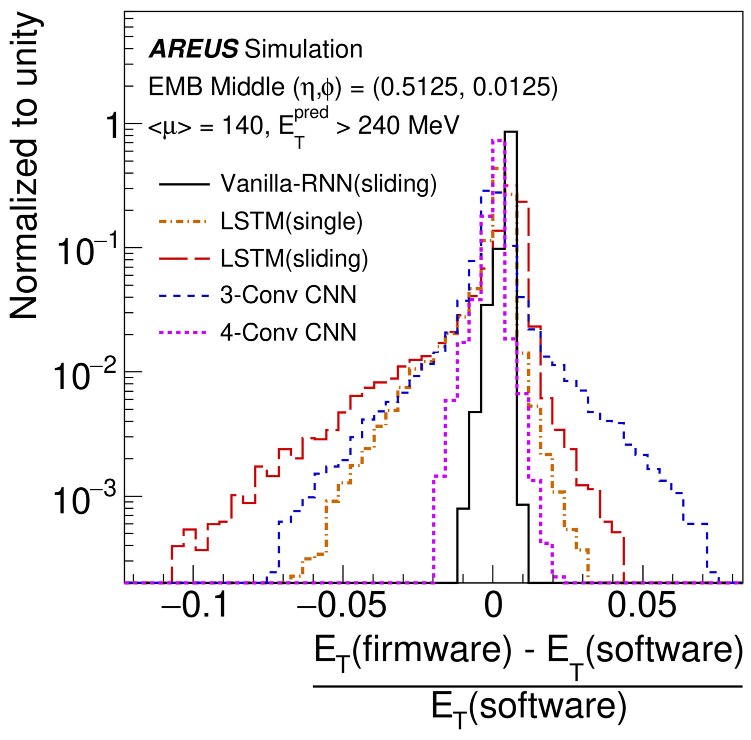


4. FPGA Implementation and Performance

- CNN: Direct **implementation in VHDL**
  - Optimal use of DSP resources on FPGA
  - Architecture automatically configured from files obtained during training
- RNN: **High Level Synthesis** approach
  - Easier design and more flexibility in architecture

- **Good agreement** between **firmware simulation** results and **software reference model**
- Inherent inaccuracies from fixed point numbers

- Performance and resource usage for single instance (left) and multiplexed model (right):



	3-Conv CNN	4-Conv CNN	Vanilla RNN (sliding)	LSTM (single)	LSTM (sliding)
Frequency F_{max} [MHz]	493	480	641	560	517
Latency clk_{core} cycles	62	58	206	220	363
#DSPs	46 0.8%	42 0.7%	34 0.6%	176 3.1%	738 12.8%
#ALMs	5684 0.6%	5702 0.6%	13115 1.4%	18079 1.9%	69892 7.5%

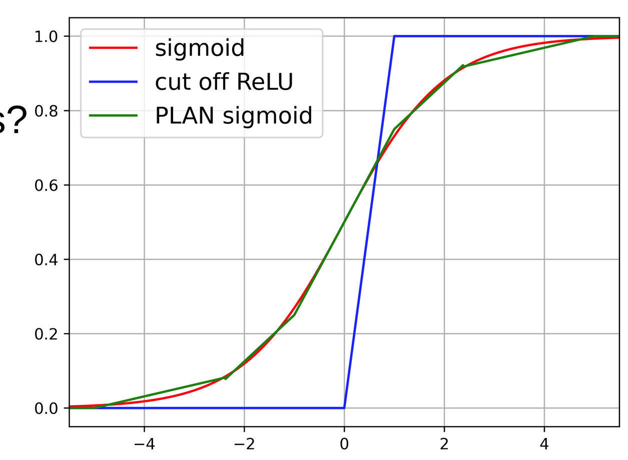
	3-Conv CNN	4-Conv CNN	Vanilla RNN
Multiplicity	6	6	15
Frequency F_{max} [MHz]	344	334	640
Latency clk_{core} cycles	81	62	120
Max. Channels	390	352	576
#DSPs	46 0.8%	42 0.7%	152 2.6%
#ALMs	14235 1.5%	15627 1.7%	5782 0.6%

- CNNs:
  - short latency and small usage of Digital Signal Processors (DSP)
  - Usage of FPGA logic (ALM) and maximum execution frequency of multiplexed version need to be optimized further
- Vanilla RNN:
  - meets requirements for resource usage and maximum clock frequency in mutltiplexed version

- **Energy reconstruction using CNNs/RNNs can be implemented on LASP FPGA, shows good agreement between software and firmware model and outperforms OF**

5. Outlook

- Reliability of ANNs for **varying pulse shapes** and sequences must be tested in more detail:
  - Influence of bunch train structure expected at HL-LHC
  - Temporal and spatial variation of proton-proton collisions at same bunch crossing leading to shifts of pulse digitization
  - Influence of varying pulse shapes in different detector regions
    → Can the same network structure be used in all detector regions?
- Studies on **reproducibility of performance** for multiple trainings of same ANN architecture:
  - How often do we have to retrain / recalibrate the ANNs?
  - Training is a statistical process
    → How many trainings are needed to obtain best performance?
  - Can optimized loss functions increase the
  - reproducibility?
- All ANN architectures must be **optimized for minimal FPGA resource usage**
- **FPGA implementation is ongoing**:
  - Multiplexing of CNNs and RNNs
  - Optimization of internal fixed point calculation



Examples for resource-saving replacements for sigmoid activation function

6. References

1. The ATLAS Collaboration, ATLAS Liquid Argon Calorimeter Phase-II Upgrade: Technical Design Report, CERN-LHCC-2017-018, ATLAS-TDR-027, CERN, 2017
2. The ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 S08003, 2008
3. W.E. Cleland, E.G. Stern, Signal processing considerations for liquid ionization calorimeters in a high rate environment, NIM A Vol. 338, 467-497, 1994
4. G. Aad, et al. , Computing and Software for Big Science 5, 19 (2021) https://link.springer.com/article/10.1007/s41781-021-00066-y

Christian Gutsche on behalf of the ATLAS Liquid Argon Calorimeter group
Institute of Nuclear and Particle Physics, TU Dresden, Germany

TECHNISCHE UNIVERSITÄT DRESDEN
INSTITUTE OF NUCLEAR AND PARTICLE PHYSICS
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung