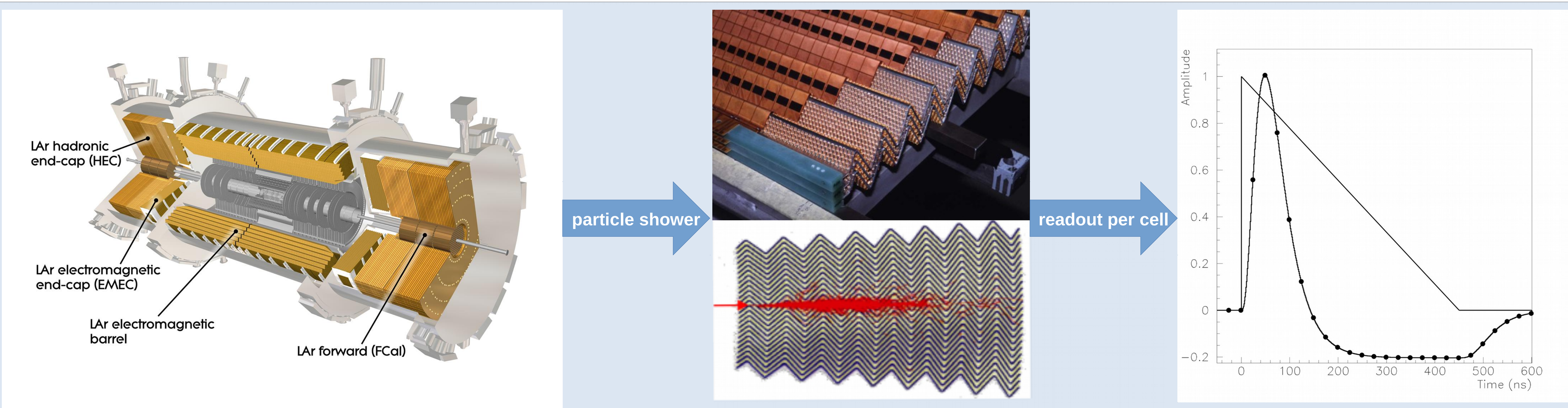
2021 International Conference on Technology and Instrumentation in Particle Physics

Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters

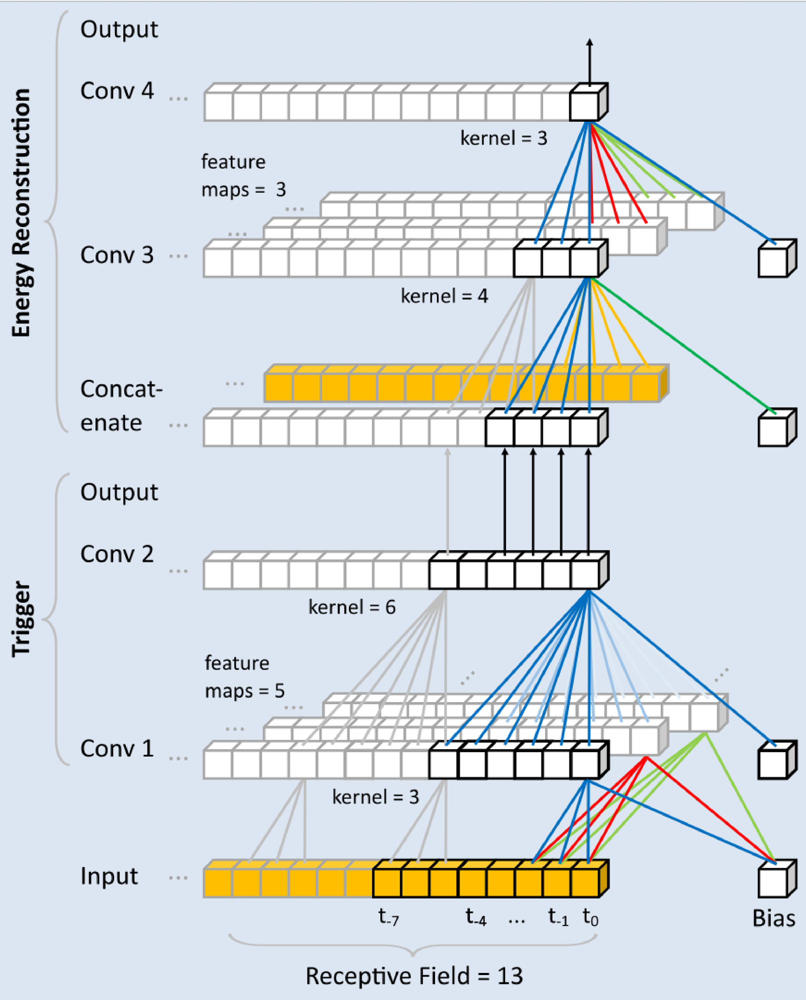
ATLAS Liquid Argon Calorimeters

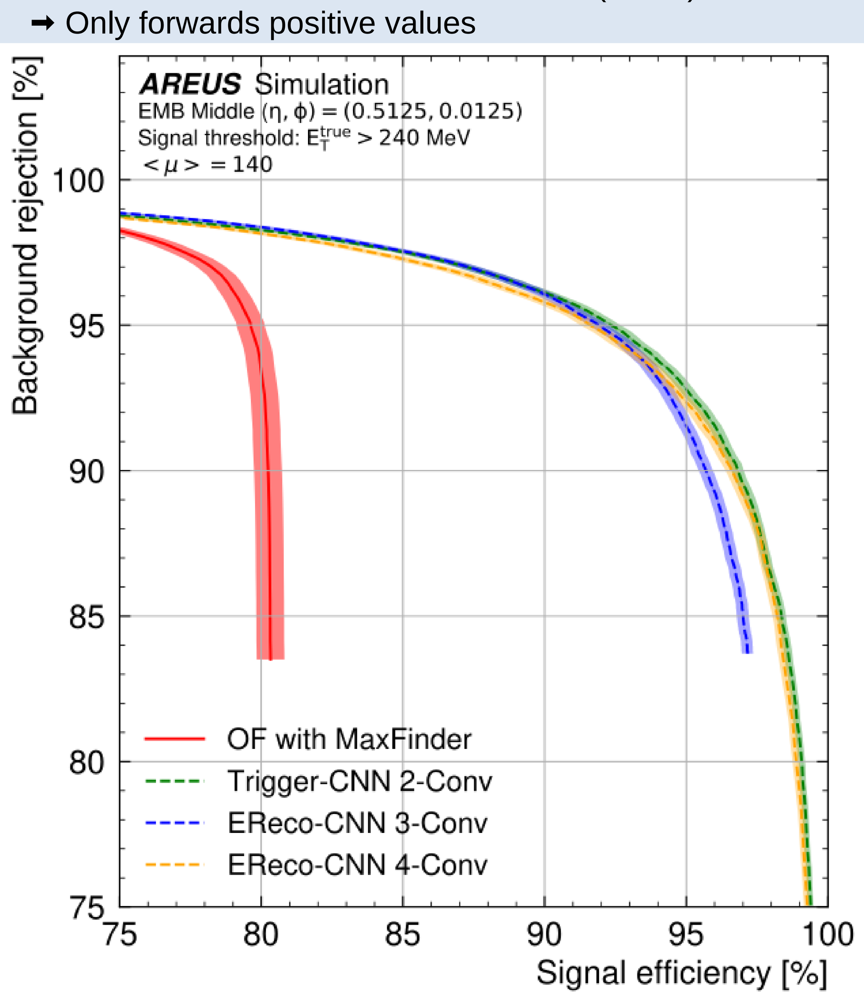


- Sampling calorimeter with ~180k cells for measuring energy deposits of electrons, photons and jets
- Triangular ionization pulse is amplified, shaped and sampled at 40 MHz

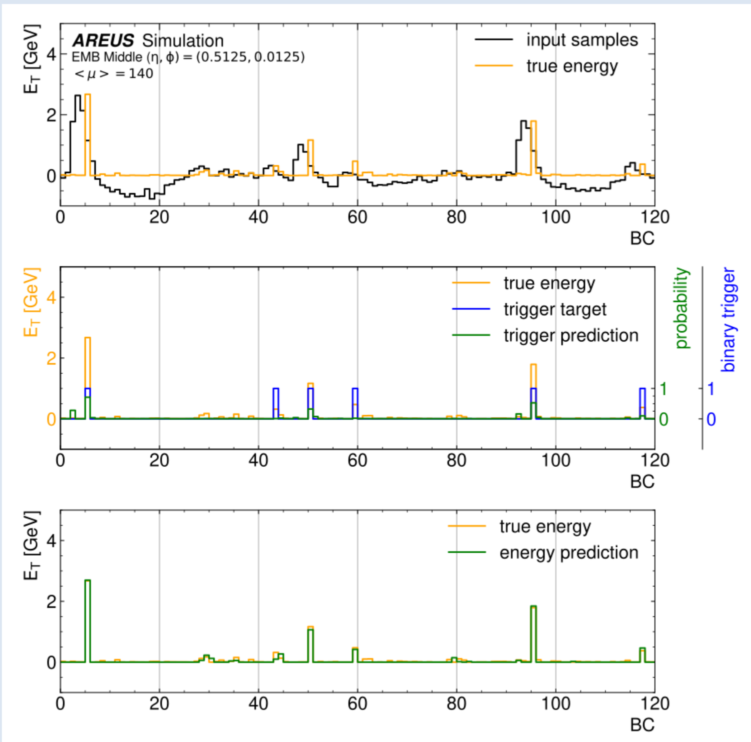
Convolutional Neural Networks (CNNs)

- Linear combination of subsequent samples as for currently used optimal filter (OF), but with more hyper-parameters, like layers, feature maps and activation functions
- **Trigger** sub-network detecting energy deposits over electronic noise threshold
- Pre-training of trigger part increases performance
- **Energy Reconstruction** sub-network uses trigger output and raw ADC samples to calculate energy
- Activation function: Rectified Linear Unit (ReLU)
  ➡ Only forwards positive values



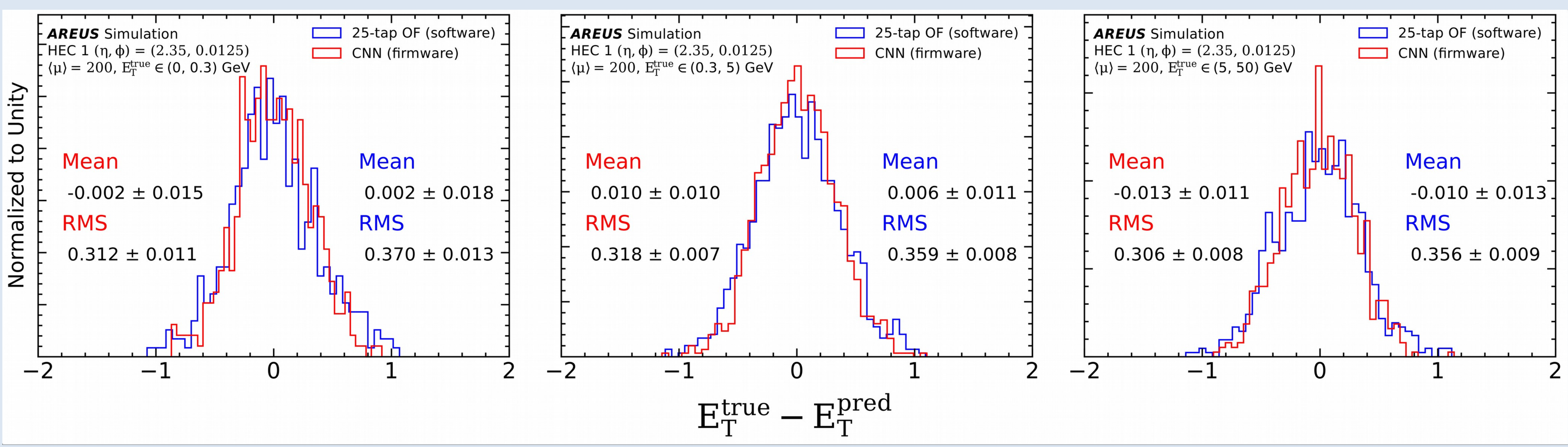


➡ Increased signal efficiency and background rejection compared to OF for trigger as well as combined network



		Layer	Kernel Size	Dilation Rate	Feature Maps	Activation Function	Number of Parameters	Receptive Field
"3-Conv"	Trigger	1	3	1	5	sigmoid	51	28
		2	6	1	1	sigmoid		
	Energy Re-construction	3	21	1	1	ReLU	43	
"4-Conv"	Trigger	1	3	1	5	sigmoid	51	13
		2	6	1	1	sigmoid		
	Energy Re-construction	3	4	1	3	ReLU	37	
		4	3	1	1	ReLU		

Energy range dependent deviation of energy resolution ➡ Additional penalty term in loss function:



$E_T^{true} - E_T^{pred}$

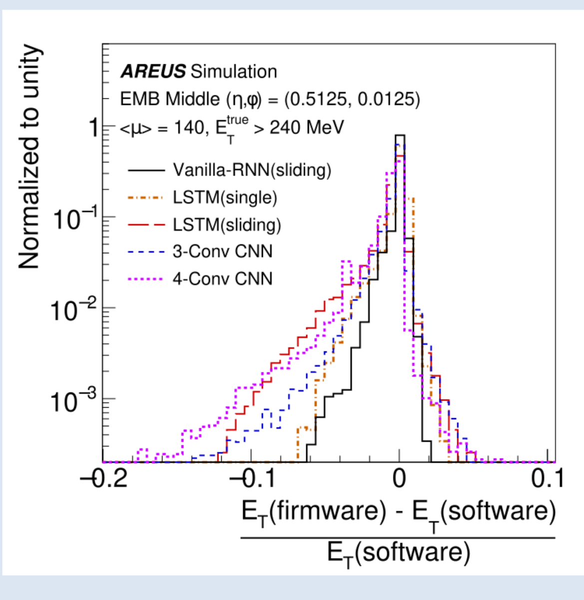
FPGA Implementation

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

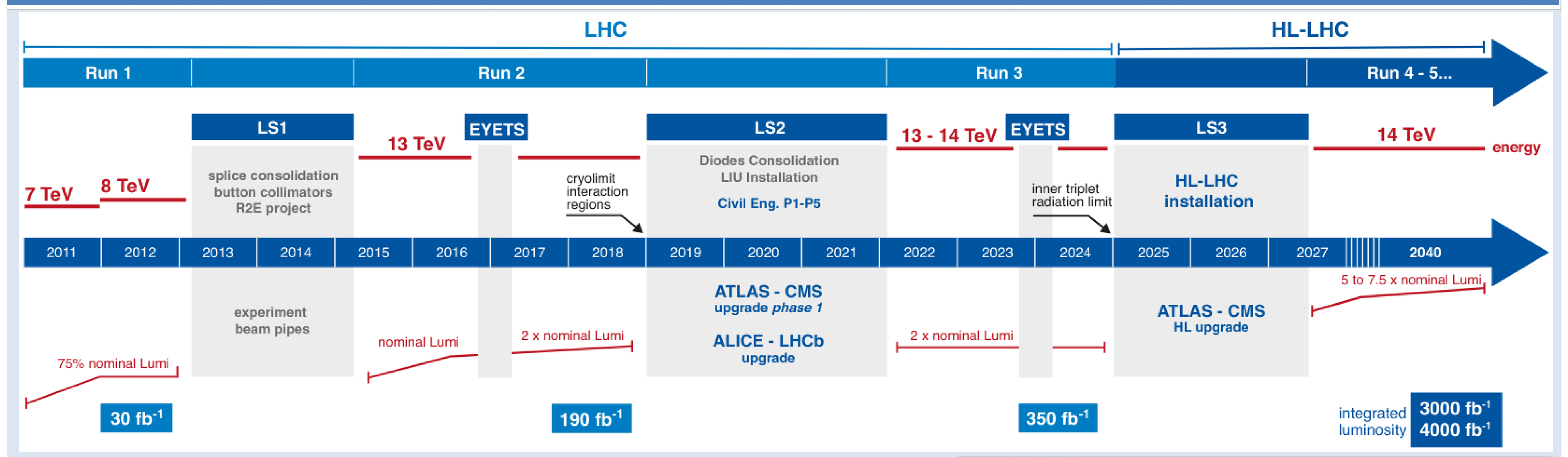
Direct **VHDL** implementation for CNNs
- Optimal usage of DSPs on the FPGA
- Modular firmware design adopting to model files from training

High Level Synthesis for RNNs
- Additional design flexibility
- One LSTM cell instance for single-cell, five for sliding-window implementation

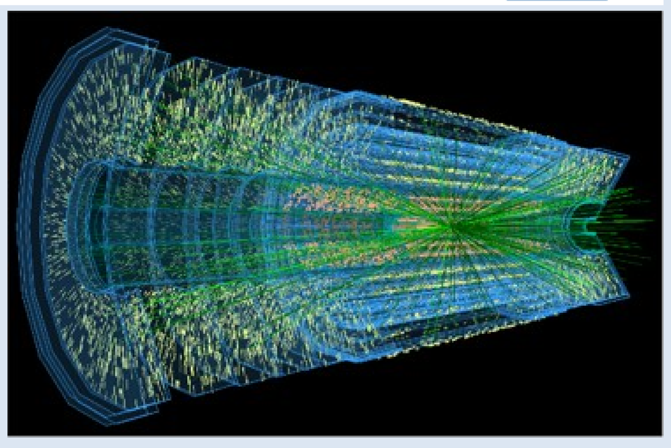
➡ CNNs use less resources allowing the processing of more channels per FPGA
➡ LSTMs candidates for readout processing with less stringent latency constraints



Phase-II Readout Electronics Upgrade

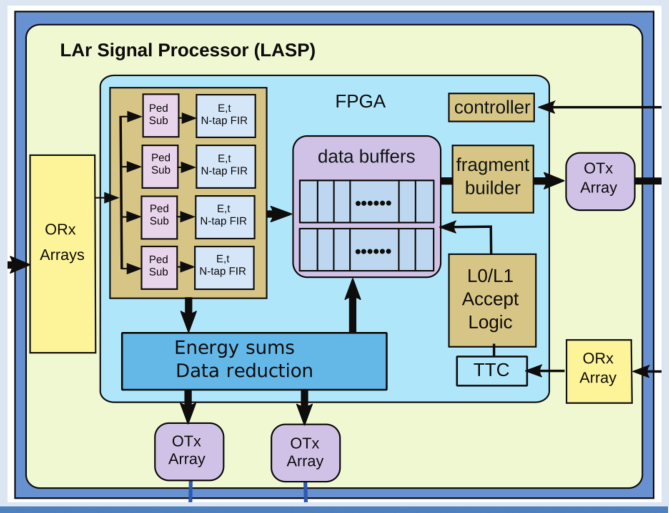


2027: High Luminosity phase of **Large Hadron Collider** (HL-LHC) starts
- Expected luminosities of up to 7.5 times the nominal value
- Mean of up to 200 simultaneous proton-proton collisions



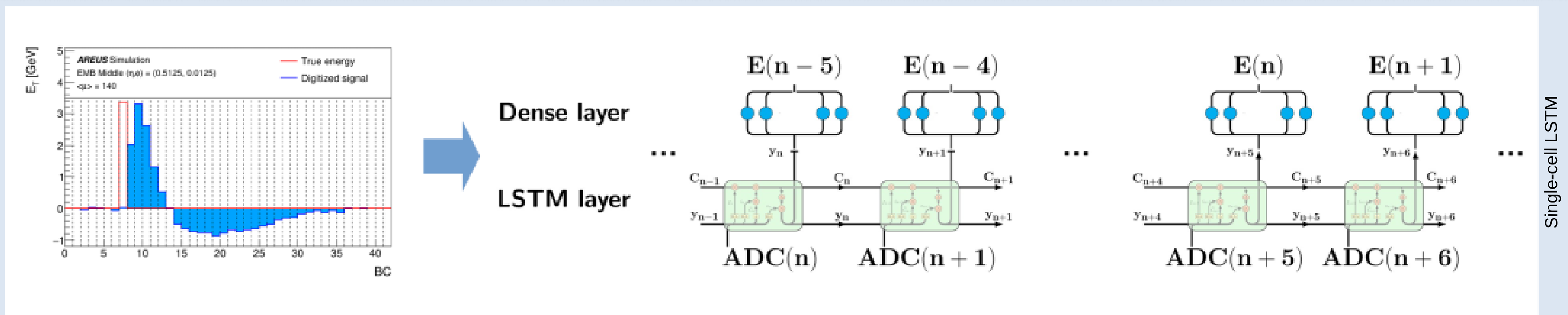
Challenges for the **LAr calorimeters**:
- Overlap of up to 25 signal pulses created in subsequent bunch crossings possible
- New trigger scheme allowing selection of events in subsequent bunch crossings

Installation of new LAr Signal Processor (LASP) boards in so-called **Phase-II Upgrade** during Long Shutdown 3
- FPGA for implementation of advanced real-time energy reconstruction algorithms
- Maximum latency of about 150 ns for energy reconstruction algorithm
- 512 LAr calorimeter cells to be processed by one FPGA



Simulations performed with dedicated ATLAS Readout Electronics Simulation framework **AREUS**
- Generates digitized pulse sequences
- Takes analog and digital electronic noise, as well as LHC bunch patterns into account

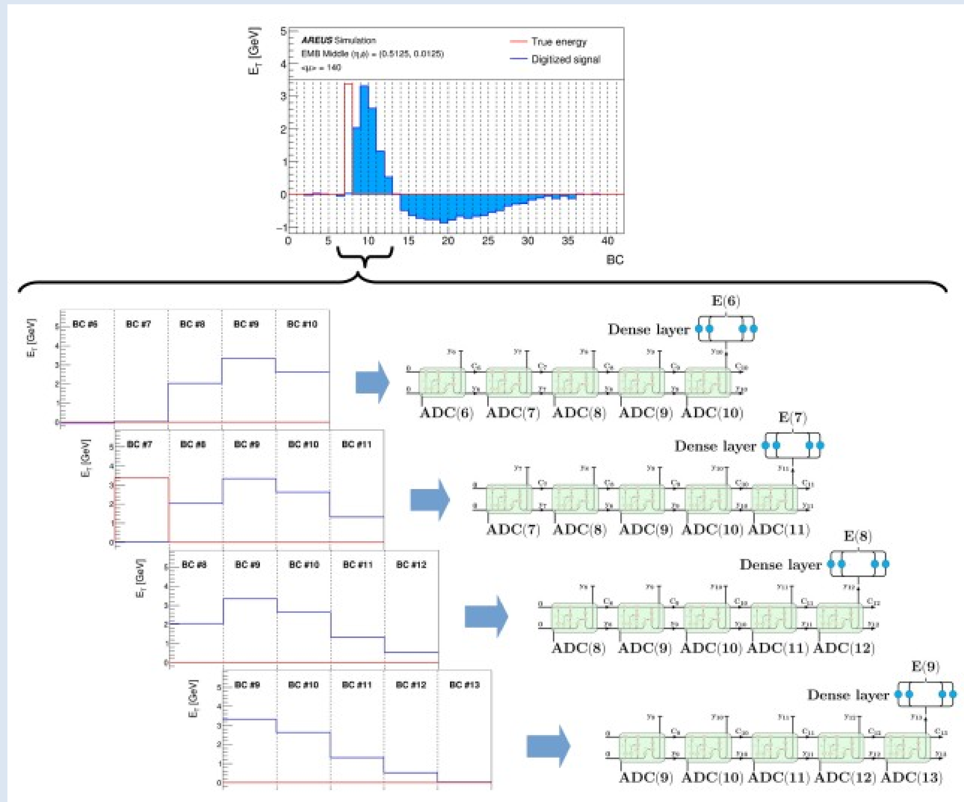
Recurrent Neural Networks (RNNs)



- Designed for inference of time series and extraction of underlying parameters
  ➡ Applies to LAr energy reconstruction
- **Long Short-Term Memory (LSTM)** architectures optimal for long sequences
- Restrictions on internal network dimensions of LSTM cells and limitation on one layer to meet FPGA resource constraints
  ➡ Vanilla-RNN with less parameters and lower expected size on hardware
- Single-neuron dense layer for decoding LSTM output and calculating the energy

Single-cell application (many-to-many RNN)
- Same operation repeated until the end of data
- Expected higher robustness for overlapping pulses

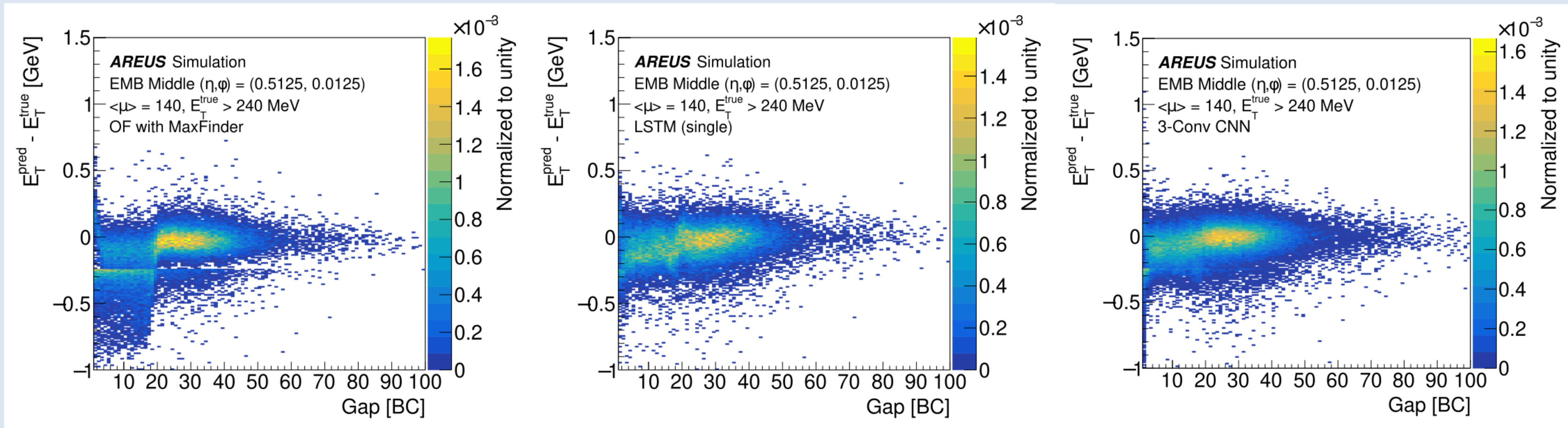
Sliding-window application (many-to-one RNN)
- Focus on few subsequent bunch crossings of interest
- Expected higher robustness for isolated data pulses



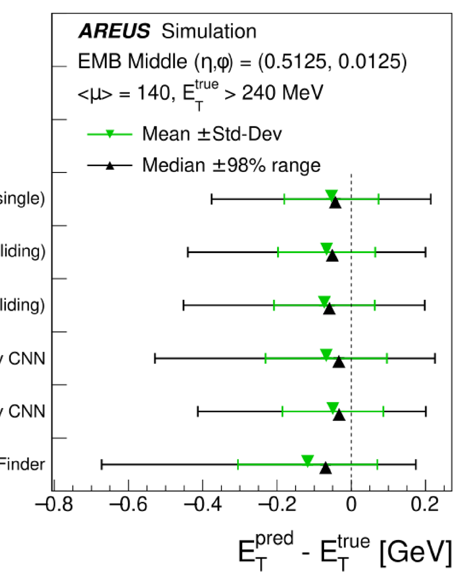
Sliding-window LSTM

Type	Time inference Receptive Field	Delay	LSTM layer Dimension	Activation	Recurrent Activation	Dense layer Dimension	Activation	Number of Parameters
Single-cell	∞	6	10	tanh	sigmoid	1	ReLU	491
Sliding-window	5	5	10	tanh	sigmoid	1	ReLU	491

Performance



- CNN and LSTM networks outperform OF in terms of bias in mean and of resolution
- Artificial neural network algorithms are robust against pulse shape distortion by overlapping events
  ➡ Improved energy reconstruction at small time gaps



References:
1. Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter, Tech. Rep. CERN-LHCC-2017-018, CERN, Geneva, Sep, 2017. https://cds.cern.ch/record/2285582.
2. W. Cleland and E. Stern, Signal Processing Considerations for Liquid Ionization Calorimeters in a High Rate Environment, Nucl. Inst. Meth. A 338 (1994) no. 2, 467–497.

Nick Fritzsche (TU Dresden) on behalf of the ATLAS Liquid Argon Calorimeter group