

Status and plans at TU Dresden

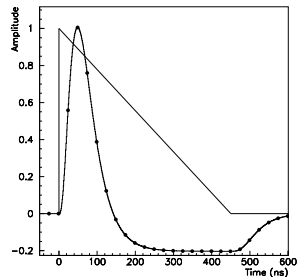
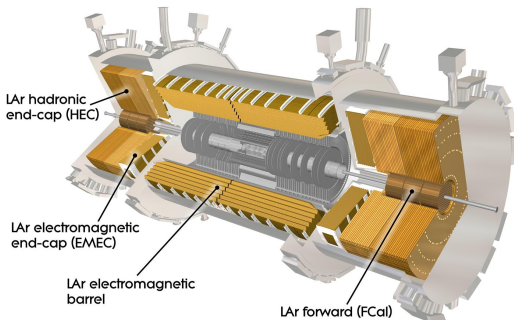
TA5-WP2: Working meeting - neural networks on FPGAs

Johann Voigt – TU Dresden

20 June 2022



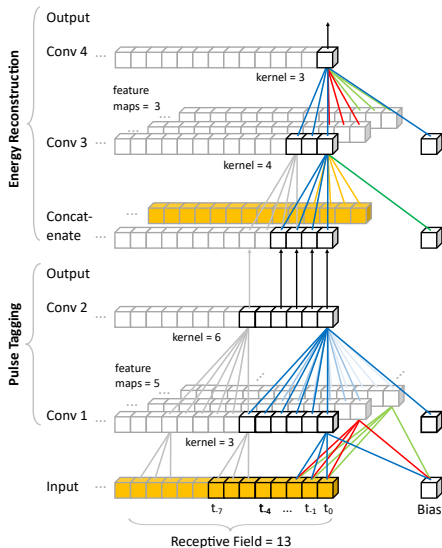
LAr-Calorimeter



- Triangular detector pulses → Analogue pulse shaping → Digitization
- Digital energy reconstruction with Optimal Filter (OF) + maximum finder for trigger

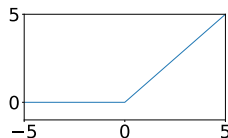
$$E(t) = \sum_i c_i \cdot x(t - i)$$

CNN architecture for energy reconstruction

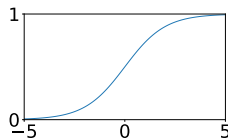


- Input: 1D time series of ADC samples (one detector cell)
- Energy reconstruction subnetwork:

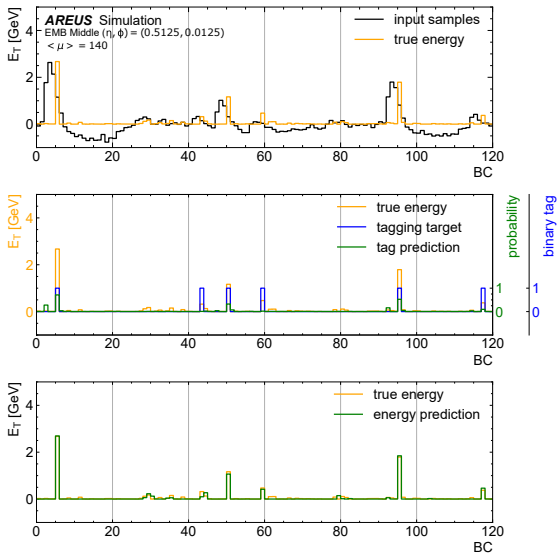
- ReLU activation function



- Tagging subnetwork:
- Sigmoid activation function

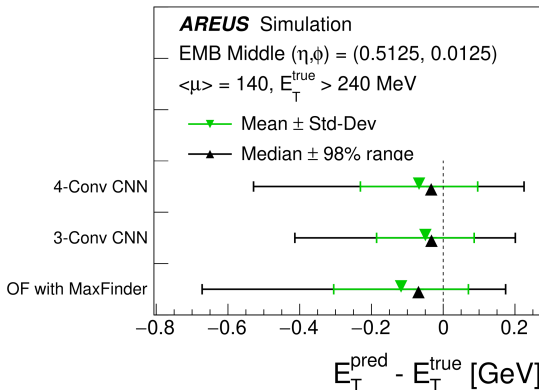


CNN example sequence



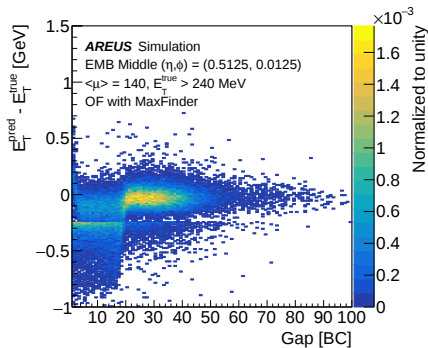
- Input sequence from AREUS simulation
- Tagging sub-network trained to output detection probability based on binary training target (240 MeV threshold)
- Energy reconstruction with true hit energy as target

Energy resolution

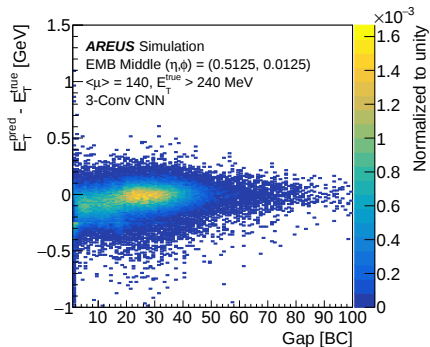


- CNNs show better energy resolution and less bias than OF
- 3-Conv is best performing CNN

CNN energy resolution as a function of gap



Optimal Filter

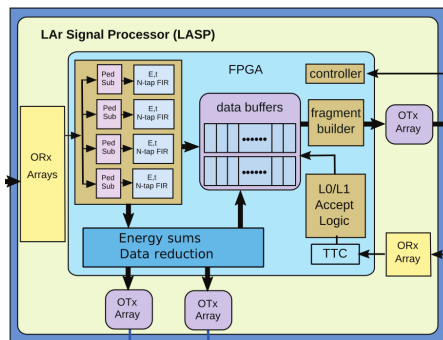


3-Conv CNN

→ Significant improvement in reconstruction of overlapping pulses

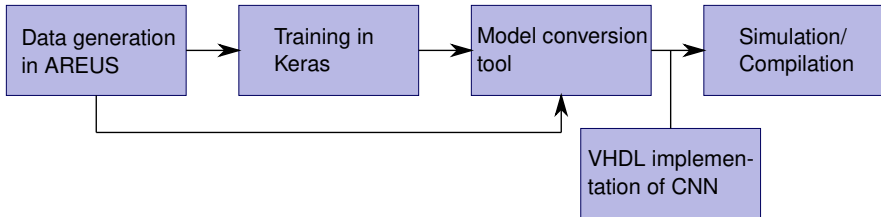
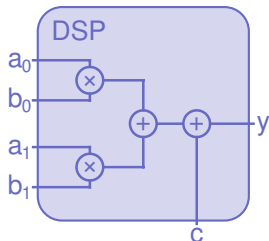
Requirements for use in off-detector electronics

- Latency below ≈ 150 ns
- Implementation on Intel Agilex FPGA (formerly Stratix 10)
 - 384 input channels @40 MHz per FPGA
 - Limited resources (DSPs and logic cells (ALMs))
- Integration with rest of readout firmware: Liquid Argon Signal Processor (LASP)

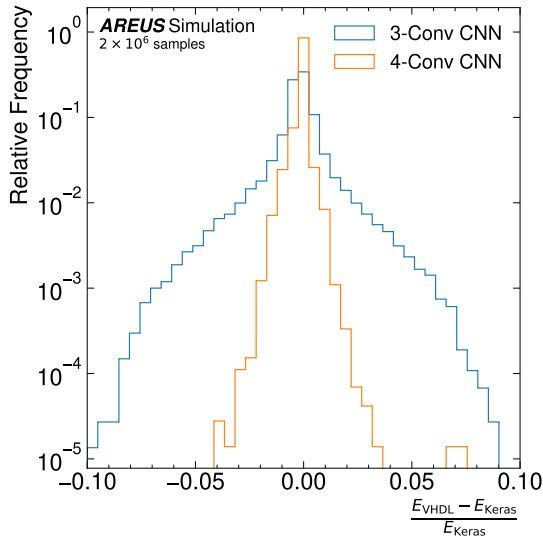


CNN firmware implementation

- Flexible CNN model implemented directly in VHDL
- Lookup table required for sigmoid activation function
- Optimized for DSP usage and latency
- DSPs can be chained for efficient multiply-add structures
- Depends on special architecture of Stratix 10 DSPs
- Fixed point calculation with 18 bit total bit width



Relative deviation between firmware and software



- Only samples with predicted energy over 240 MeV included
- Good agreement between firmware and software
- Inherent deviations due to fixed point calculation
→ Potential problems with very high/low weights

Compilation results for multiplexing

	4-Conv 1x	4-Conv 8x	4-Conv 12x	3-Conv 8x	3-Conv 12x
$f_{\max}$	432 MHz	377 MHz	346 MHz	387 MHz	351 MHz
ALMs	5473	15988	18453	16077	20107

- DSP usage independent of multiplexing (3-Conv: 46, 4-Conv: 42)
- ALM usage still needs optimization
- Latency for 4-Conv with $12\times$ multiplexing: 72 clock cycles (= 150 ns if performance is optimized to run at targeted 480 MHz)

Summary

- Flexible VHDL implementation supporting CNNs with dilation and input concatenate layers for 1D continuous input stream
- Good agreement between CNN firmware implementations and software (Keras and fixed-point reference model)
- Currently depends on project specific framework
- Only runs on Intel Stratix 10 (and similar Intel FPGAs)
- Maximum clock frequency and ALM usage in multiplexed version need further optimization
- Further information about training and performance available in [1]

Plans for the future

- Training: More studies about robustness for slight variations in input
- Integrate with rest of readout chain and test on hardware demonstrator
- Investigate high-level synthesis options (HLS4ML) as alternative
- Tentative if CNN implementation proves useful outside of ATLAS LAr context: Split out of LASP framework and publish as open source project

Sources I

- [1] Georges Aad et al. “Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters”. 25th International Conference on Computing in High-Energy and Nuclear Physics. Geneva, Feb. 2021. URL: <https://cds.cern.ch/record/2752649>. Accepted for publication 2021.

- [2] ATLAS Collaboration. “Monitoring and data quality assessment of the ATLAS liquid argon calorimeter”. In: *JINST* 9.arXiv:1405.3768. CERN-PH-EP-2014-045 (May 2014). Plot available separately: <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/LARG-2013-01/>, P07024. 39 p. URL: <http://cds.cern.ch/record/1701107> (visited on 05/28/2017).

Sources II

- [3] Joao Pequena. *Computer generated image of the ATLAS Liquid Argon*. CERN. Mar. 27, 2008. URL: <https://cds.cern.ch/record/1095928> (visited on 03/29/2021).