

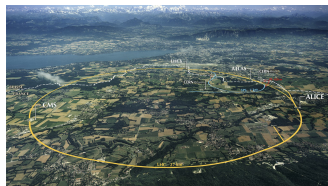
Development of an FPGA Implementation of Convolutional Neural Networks for Signal Processing for the Liquid-Argon Calorimeter at ATLAS

Nick Fritzsche

Institute of nuclear and particle physics (IKTP), TU Dresden

23 March, 2022
DPG Spring Meeting Heidelberg

The ATLAS Detector at the LHC



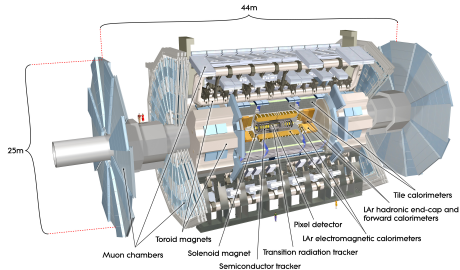
<https://cds.cern.ch/record/1295244>

The Large Hadron Collider (LHC)

- 27 km circular collider at CERN/Geneva
- achievements: Higgs-Boson, quark-gluon plasma and many more
- 25 ns spacing between proton bunches (40 MHz)

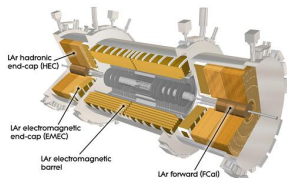
The ATLAS Detector

- inner detector with tracking system
- **calorimeters**
- muon spectrometer



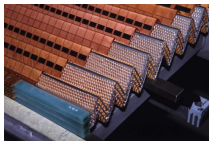
<https://cds.cern.ch/record/1095924>

Signal Readout of the ATLAS LAr Calorimeter

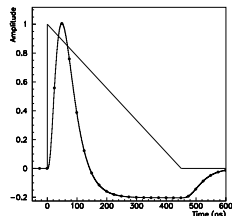
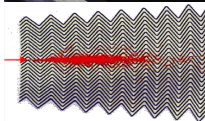


<https://cds.cern.ch/record/1095928>

particle shower



©CERN



<https://cds.cern.ch/record/331061>

readout per cell

LAr Calorimeter

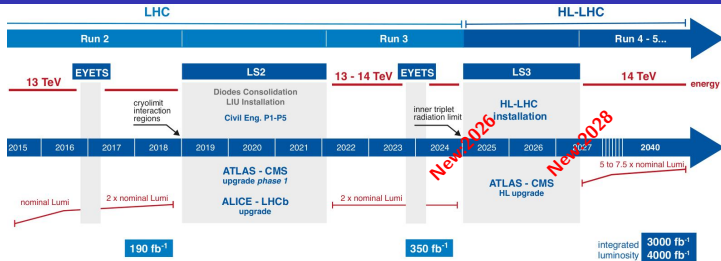
- absorber material (Pb, Cu, W) and electrodes in accordion geometry
- in between liquid argon as active medium

Signals

- energy deposits raise a triangular pulse
- shaped into bipolar pulse and digitized
- amplitude proportional to deposited energy
- energy reconstruction using Optimal Filter:

$$E_t = \sum_i w_i \cdot S_{t-i}$$

High Luminosity LHC



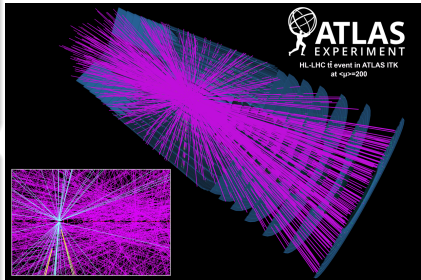
<https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>

High Luminosity Upgrade

- achieved in Run-2: 139 fb⁻¹
- goal for HL-LHC: 250 fb⁻¹ to 300 fb⁻¹ per year
→ 3000 fb⁻¹ integrated luminosity

Detector Challenges

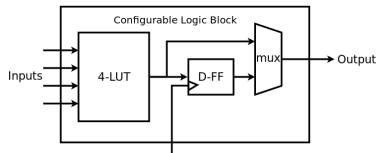
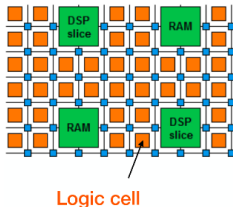
- leveled luminosity of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
- pileup of up to 200 collisions/BC
→ increased event size, trigger rate, detector occupancy, reconstruction complexity



<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/UpgradeEventDisplays>

Field Programmable Gate Arrays (FPGAs)

- integrated circuit configurable by the designer after manufacturing
- reconfigurable hardware allows testing of different firmware
- use Intel Stratix 10



<https://newsroom.intel.com/editorials/intels-stratix-10-fpga-supporting-smart-connected-revolution> <https://indico.cern.ch/event/773049/contributions/3474297>

Advantages of Implementation on FPGAs

- real time data processing with high frequencies (100 MHz – 1 GHz)
- parallel data processing
- signal processing algorithm can be reconfigured

Requirements

- signal processing algorithm should be designed such that parameters are kept at a minimum ($\approx 50-100$)
- aim core frequency for Phase II data processing of $480 \text{ MHz} = 12 \times 40 \text{ MHz}$
- minimize latency ($< 150 \text{ ns}$): results are input to trigger system
- meet resource limitations of FPGA

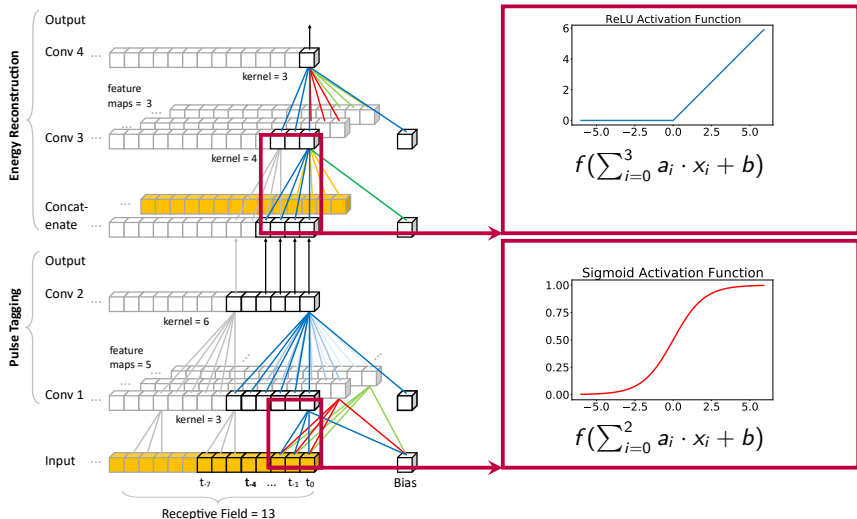
→ set pipeline registers as a compromise of the factors above

Energy Reconstruction by Convolutional Neural Networks

Our Approach

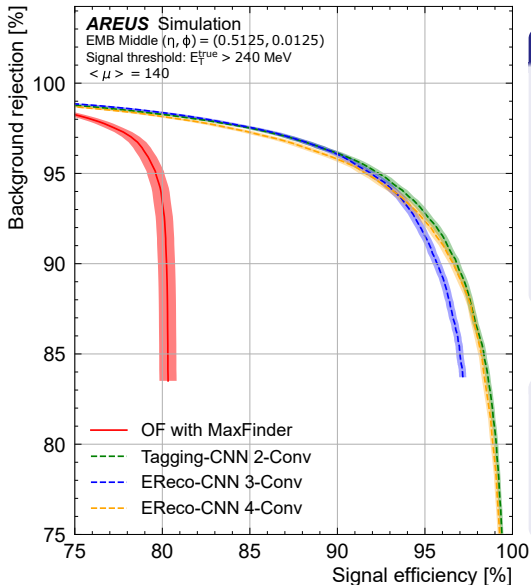
2-step convolutional neural network for energy reconstruction

<https://link.springer.com/article/10.1007/s41781-021-00066-y>



Performance Evaluation: OF vs CNN - Trigger Efficiency

<https://link.springer.com/article/10.1007/s41781-021-00066-y>



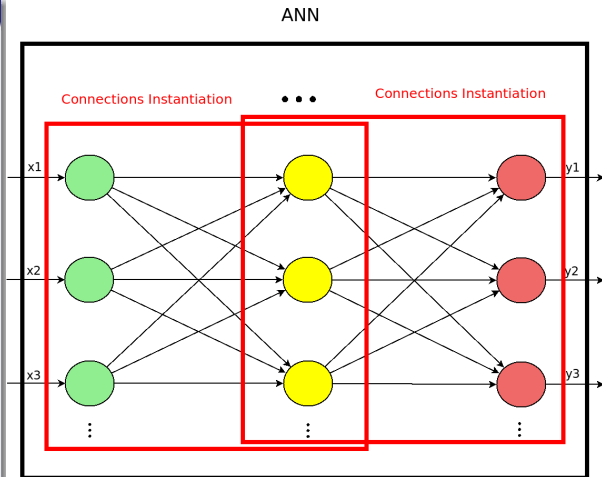
ROC curves

- indicate detection performance
- signal efficiency
$$= \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$
- background rejection
$$= \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$
- dependent on threshold

→ energy reconstruction CNNs have slightly reduced maximum efficiencies compared to their underlying Trigger CNN
→ CNNs outperform OFMax

Connections subcomponent

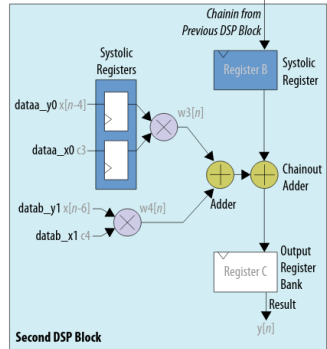
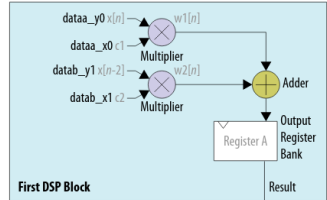
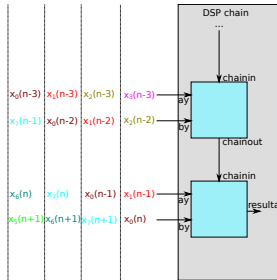
- general *connection* component for all operations between neighboring layers of ANN
 - configurable #inputs, #outputs, activation functions
- multi-layer network component chains *connection* instances
 - capable to implement different kernel sizes, #feature maps and dilated CNNs
 - configurable with file produced after training (json)



Optimization of calculations: DSP Chain

DSP Chain

- DSPs are chained up to accumulate products over whole kernel
- match timing of input ports to process data from multiple cells in one DSP chain instance
- load weights from RAM without recompilation



<https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/stratix-10/ug-s10-dsp.pdf>

Process 384 to 512 calorimeter cells per FPGA → In total 400 to 550 FPGAs needed

<https://link.springer.com/article/10.1007/s41781-021-00066-y>

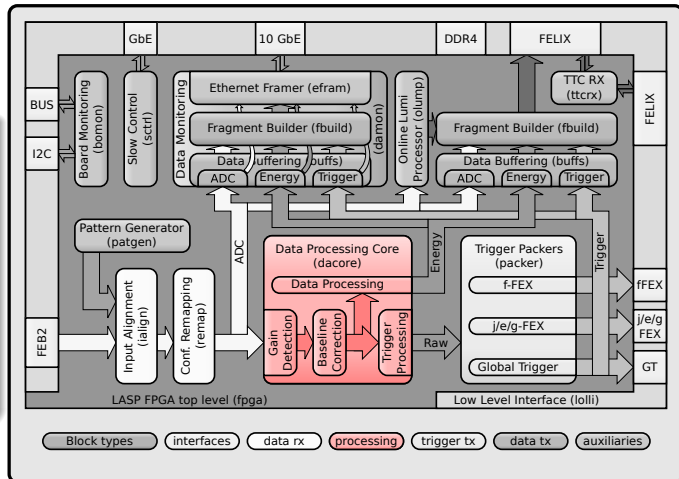
| Single Channel | | |
|--|--------------|--------------|
| | 3-Conv CNN | 4-Conv CNN |
| Frequency F_{\max} [MHz] | 493 | 480 |
| Latency clk_{core} cycles | 62 | 58 |
| Resource Usage | | |
| DSPs | 46 0.8% | 42 0.7% |
| Adaptive logic modules | 5684 0.6% | 5702 0.6% |

| Time-multiplexed | | |
|--|---------------|---------------|
| | 3-Conv CNN | 4-Conv CNN |
| Multiplicity | 6 | 6 |
| Frequency F_{\max} [MHz] | 344 | 334 |
| Latency clk_{core} cycles | 81 | 62 |
| Max. Channels | 390 | 352 |
| Resource Usage | | |
| DSPs | 46 0.8% | 42 0.7% |
| Adaptive logic modules | 14235 1.5% | 15627 1.7% |

→ Resource-efficient and short latency, but low execution frequency

Ongoing Work: Integration in LAr Signal Processor

- integrate CNNs in data processing core of LAr signal processor firmware
- output interfaces to trigger, readout and monitoring path



<https://gitlab.cern.ch/atlas-lar-be-firmware/LASP/LASP-doc/>

CNNs for Energy Reconstruction

- electronics of ATLAS LAr Calorimeters will be upgraded for HL-LHC until end of 2028
- harsh environment with up to 200 pile-up events
- energy reconstruction by convolutional neural networks shows promising results
- CNN implementation in FPGA successful
- next steps:
 - integrate CNNs in LAr signal processor data core firmware
 - perform hardware tests in test bench
 - create high level synthesis implementation and compare performance