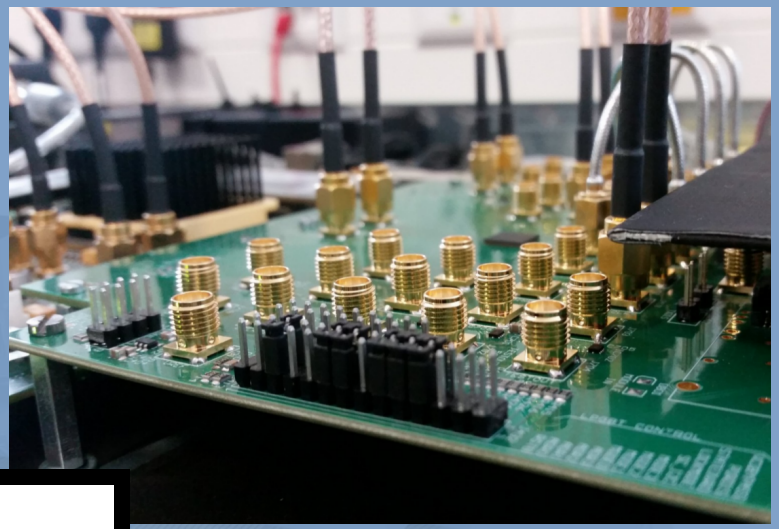
Convolutional Neural Networks for Processing of ATLAS Liquid Argon Calorimeter Signals with FPGAs

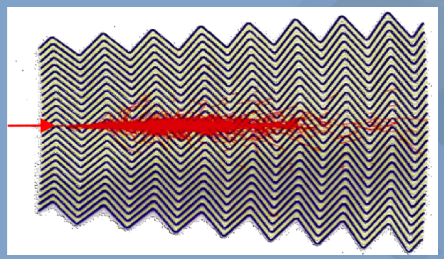
Anne-Sophie Berthold, **Nick Fritzsche**,
Markus Helbig, Rainer Hentges, Arno Straessner, Johann Christoph Voigt

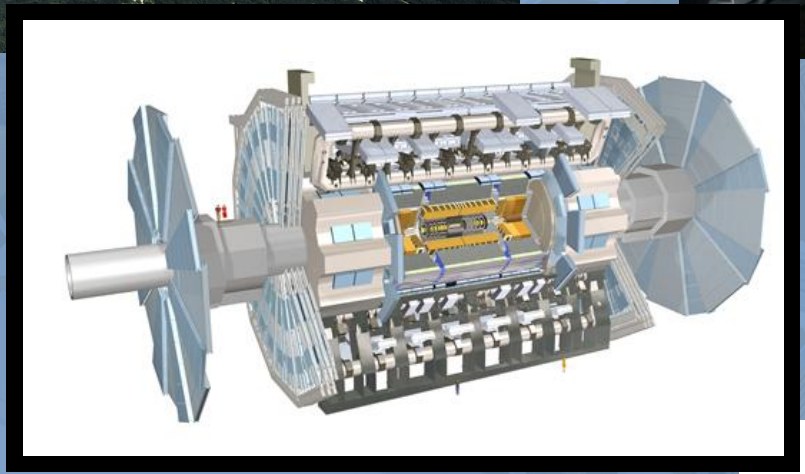
IKTP Dresden

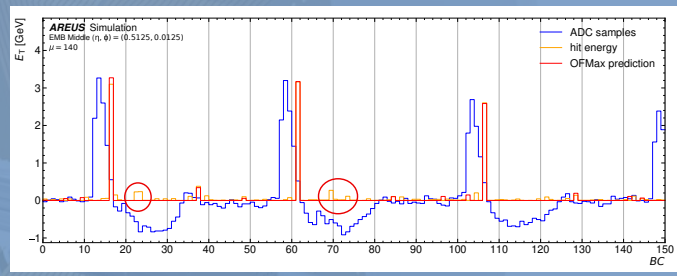
April 22, 2021

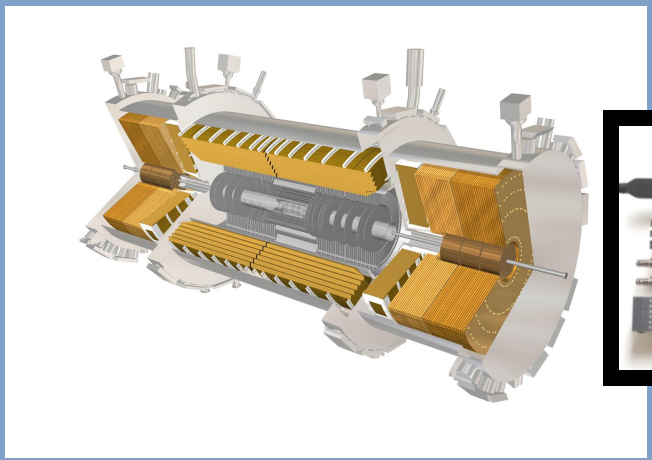

LHCb
ATLAS
CERN Meyrin
CERN Prévessin
SPS 7 km
SUISSE
FRANCE
CMS
ALICE
LHC 27 km

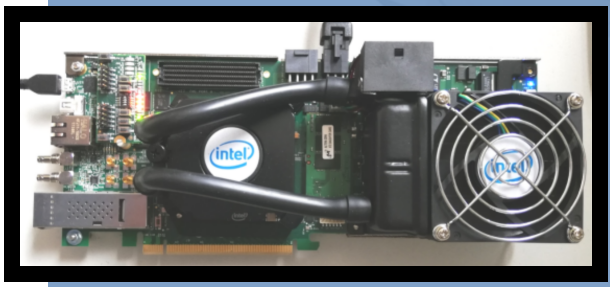


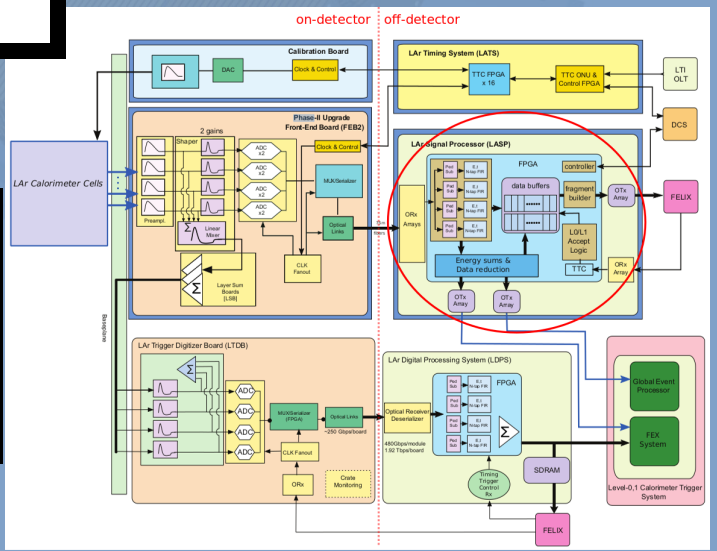




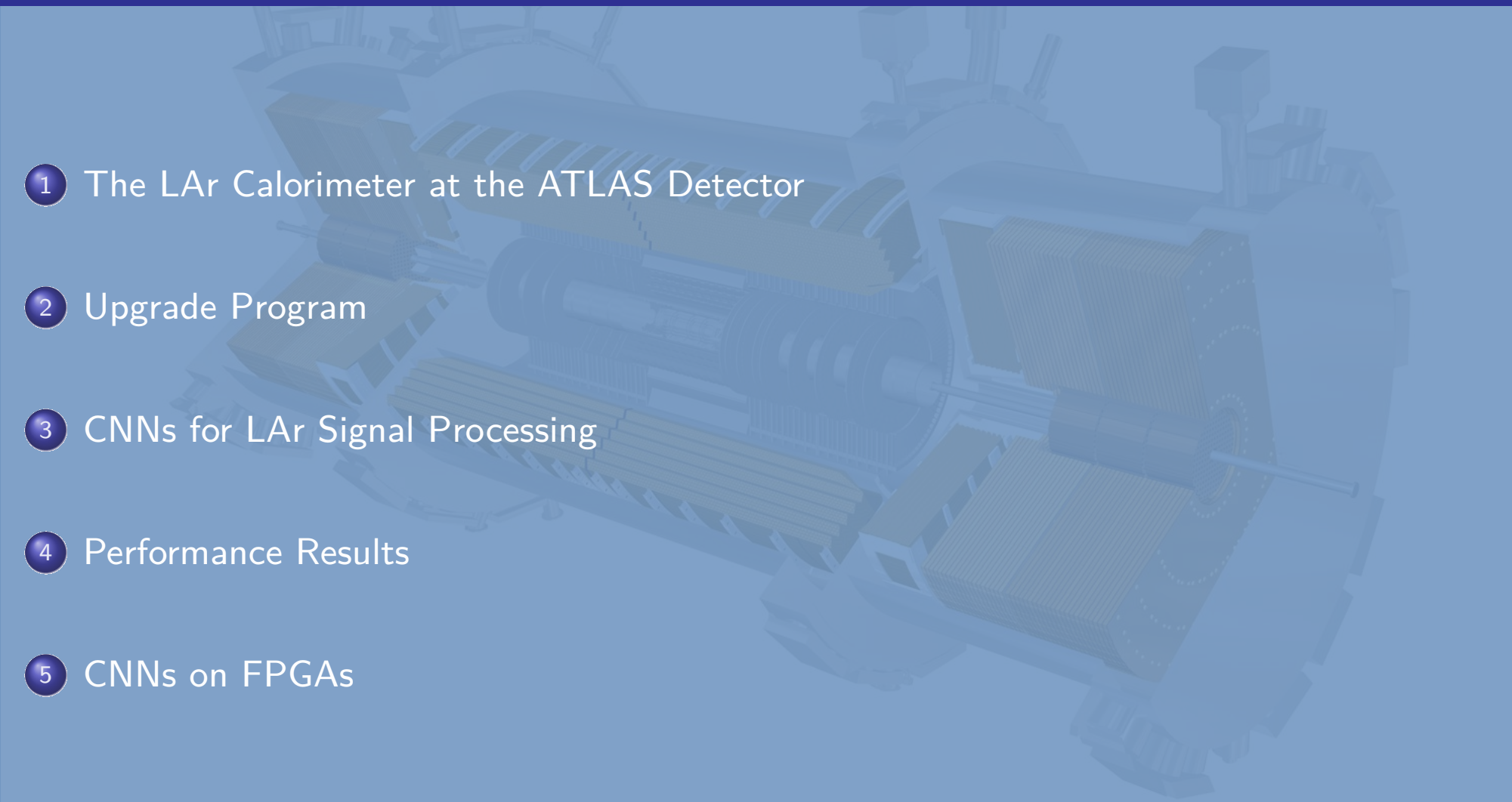

AREUS Simulation
EMB Middle (η, φ) = (0.5125, 0.0125)
μ = 140
ADC samples
hit energy
OFMax prediction
E_T [GeV]
BC






on-detector
off-detector
LAr Calorimeter Cells
Calibration Board
DAC
Clock & Control
LAr Timing System (LATS)
TTC FPGA x 16
TTC ONU & Control FPGA
LTI OLT
DCS
Phase-II Upgrade Front-End Board (FEB2)
2 gains
Shaper
Preampl.
Linear Mixer
Layer Sum Boards (LSB)
MUX/Serializer
Optical Links
CLK Fanout
LAr Signal Processor (LASP)
FPGA
controller
data buffers
fragment builder
LOL1 Accept Logic
TTC
Energy sums & Data reduction
ORx Arrays
OTx Array
FELIX
Backplane
LAr Trigger Digitizer Board (LTDB)
ADC
MUX/Serializer (FPGA)
Optical Links
~250 Gbps/board
CLK Fanout
ORx
Clock Monitoring
LAr Digital Processing System (LDPS)
Optical Receiver Deserializer
480Gbps/module
1.92 Tbps/board
FPGA
SDRAM
Timing Trigger Control for
FELIX
Global Event Processor
FEX System
Level-0,1 Calorimeter Trigger System

Overview



The ATLAS Detector at the LHC

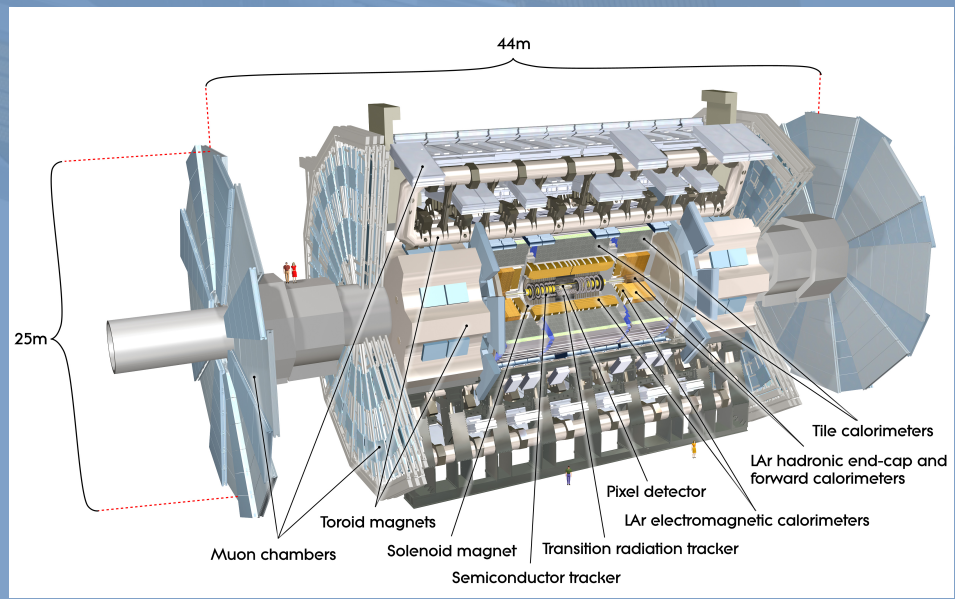
https://static1.bmbfcluster.de/3/4/3/8_ef6a5eef8f44963/3438meg_22ce2885dae52af.jpg

The Large Hadron Collider (LHC)

- 27 km circular collider at CERN/Geneva
- achievements: Higgs-Boson, quark-gluon plasma, CP-violation, increased accuracy of Standard Model parameters ...
- 25 ns spacing between proton bunches (40 MHz)

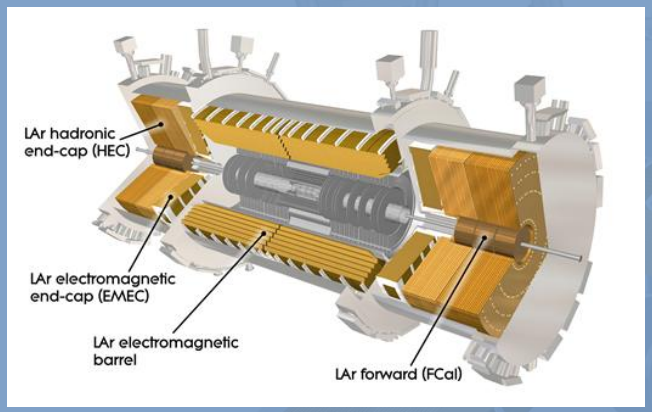
The ATLAS Detector

- three main parts:
  - inner detector with tracking system
  - **calorimeters**
  - muon spectrometer



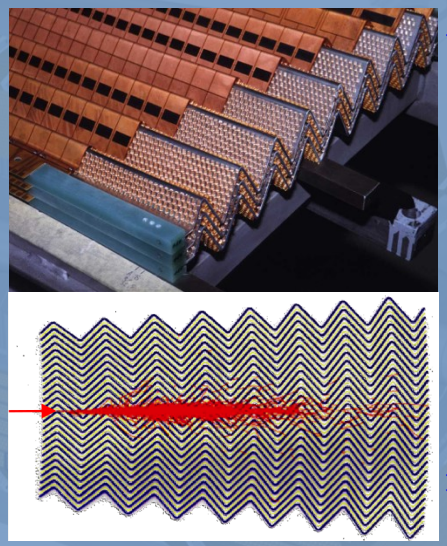
https://cds.cern.ch/record/1095924

Signal Readout of the ATLAS LAr Calorimeter



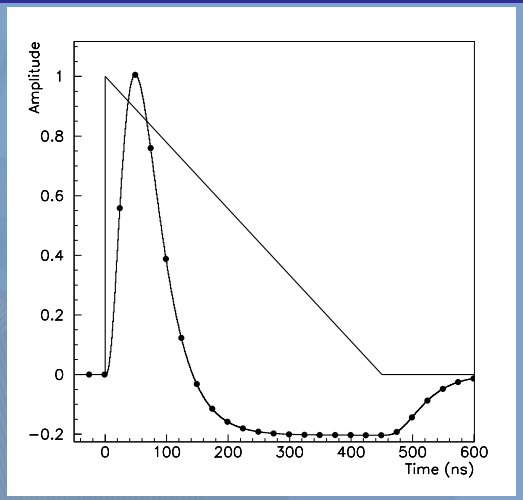
https://cds.cern.ch/record/1095928

particle shower



http://www.particles.uni-freiburg.de/dateien/vorlesungsdateien/particledetectors/kap8

readout per cell



https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003

LAr Calorimeter

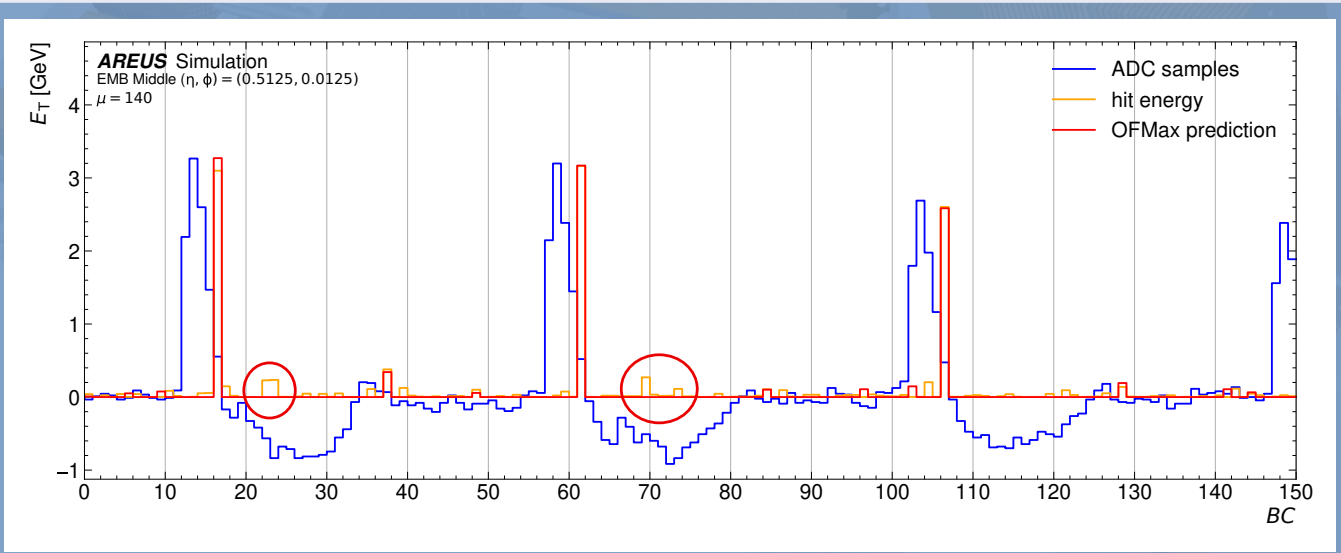
- absorber material (Pb, Cu, W) and electrodes in accordion geometry
- in between liquid argon as active medium

Signals

- energy deposits raise a triangular pulse
- shaped by CR(RC)^2 analog filter into bipolar pulse and digitized
- amplitude proportional to deposited energy

Energy Reconstruction by the **Optimal Filter**

- Optimal Filter (OF) is applied to calculate deposited energy
- it is optimized to suppress noise and reconstruct peak timing
- the trigger system applies an additional maximum finder, which is potentially insensitive during the undershoot of a pulse



$$y(n) = \sum_{k=0}^{M-1} a(k)x(n-k)$$

y ... OF output

a ... OF coefficients

x ... ADC samples

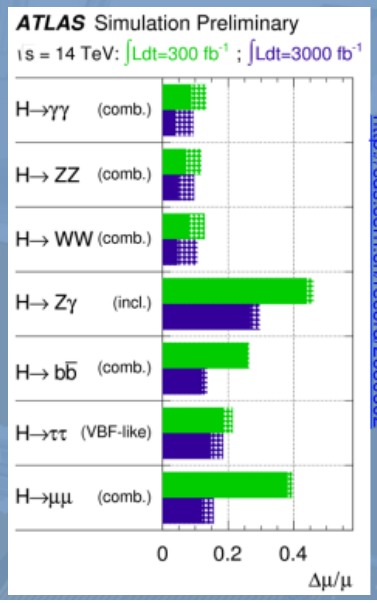
M ... OF filter depth

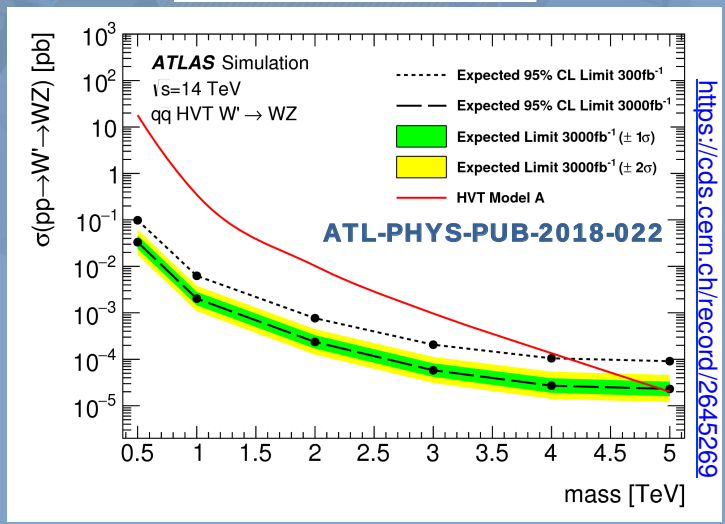
The current system is not enough ...

Physics Motivation

upgrade LHC and ATLAS for better physics performance:

1. precision measurements: Investigate Higgs coupling and further SM processes
   - SM rare decays, like H $\rightarrow \mu\mu$
   - Higgs self-coupling
   - Higgs boson couplings
2. search for Beyond Standard Model physics (BSM):
   - SUSY
   - dark sector
   - long lived particles





High Luminosity LHC

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

High Luminosity Upgrade

- achieved in Run-2: $135\,\mathrm{fb}^{-1}$
- expectation for Run-3: $150\,\mathrm{fb}^{-1}$
- goal for HL-LHC: $250\,\mathrm{fb}^{-1}$ to $300\,\mathrm{fb}^{-1}$ per year $\longrightarrow$ $3000\,\mathrm{fb}^{-1}$ integrated luminosity

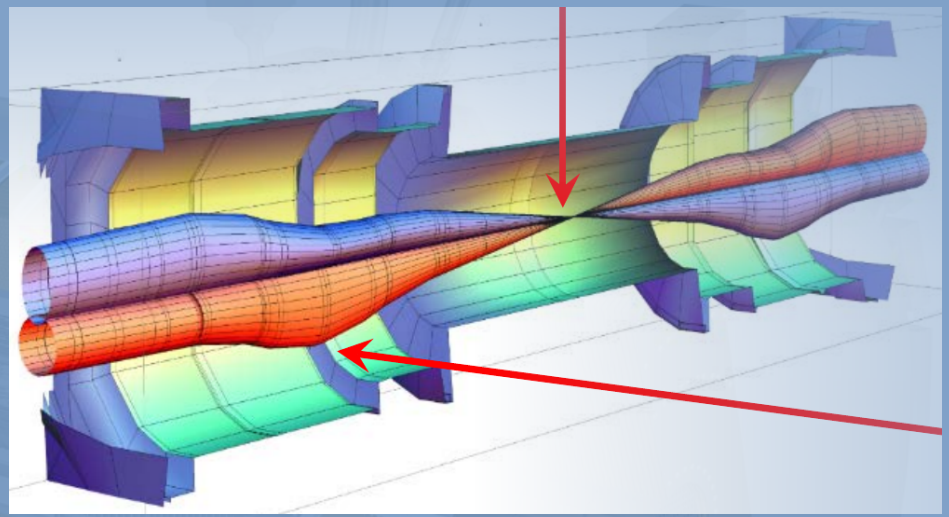
High Luminosity LHC

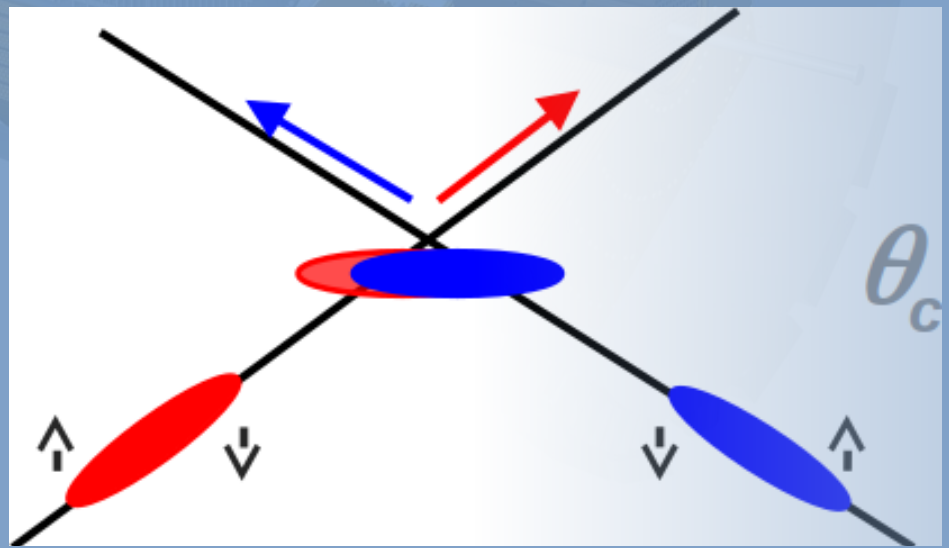
LHC Changes

- improved injectors (Linac-4)
  $\longrightarrow$ more protons/bunch, brighter
- rotate bunches in crab cavities to maximise bunch overlap in collision regions

Detector Challenges

- leveled luminosity of
  $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
- pileup of up to 200 collisions/bunch crossing
  $\longrightarrow$ increased event size, trigger rate, detector occupancy, reconstruction complexity
- radiation levels of up to 10 MGy
  $\longrightarrow$ increased radiation damage and activation of materials





http://cds.cern.ch/record/2642101/files/ATL-PHYS-SLIDE-2018-841.pdf

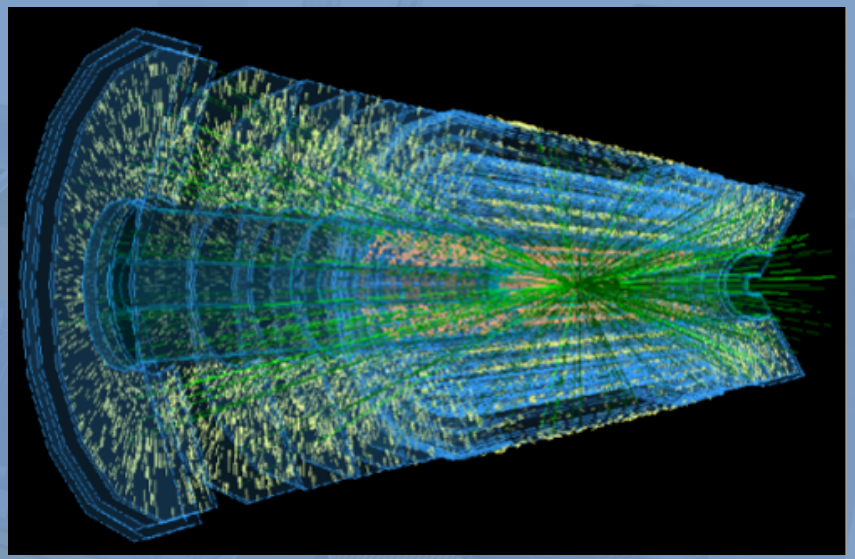
High Luminosity LHC

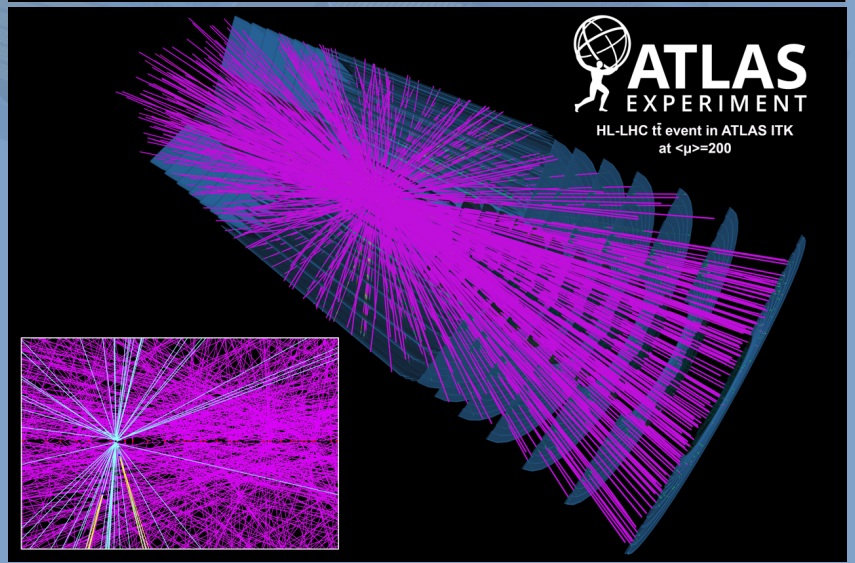
LHC Changes

- improved injectors (Linac-4)
  $\longrightarrow$ more protons/bunch, brighter
- rotate bunches in crab cavities to maximise bunch overlap in collision regions

Detector Challenges

- leveled luminosity of
  $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
- pileup of up to 200 collisions/bunch crossing
  $\longrightarrow$ increased event size, trigger rate, detector occupancy, reconstruction complexity
- radiation levels of up to 10 MGy
  $\longrightarrow$ increased radiation damage and activation of materials





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

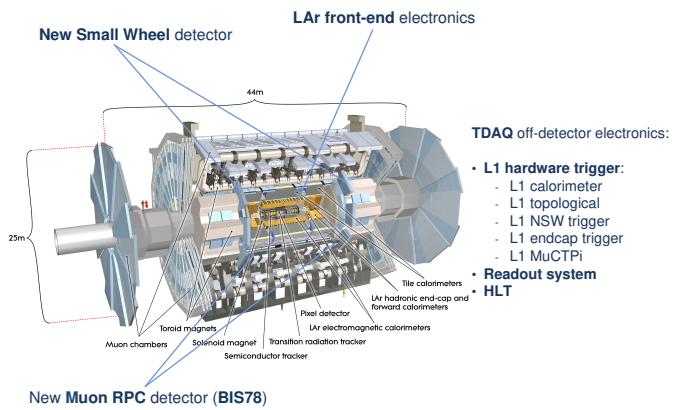
ATLAS Phase-I and Phase-II Upgrades

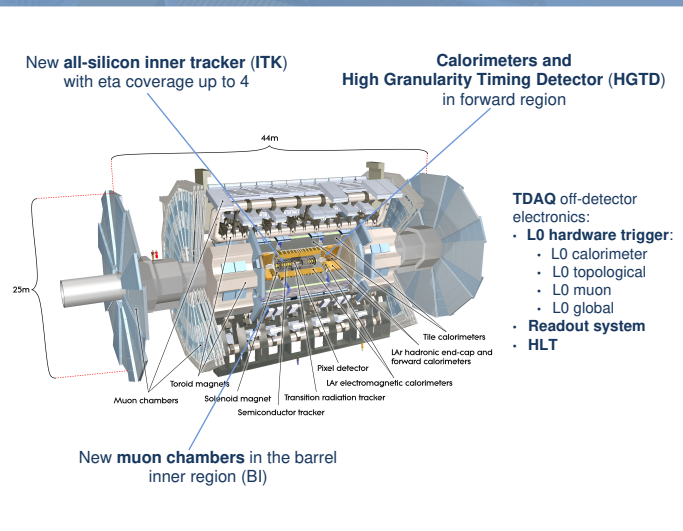
Phase-I Upgrades

- improved rate capability and background rejection
- installation is ongoing

Phase-II Upgrades

- new tracker and timing detector for improved pile-up handling
- upgrades on muon and trigger system increase trigger and readout capabilities
- trigger rate increases from 100 kHz to 1 MHz
- longer time until trigger decision (latency): 30 μs instead of 1.5 μs
- tracker must nevertheless be read out after $\approx 3\,\mu$s



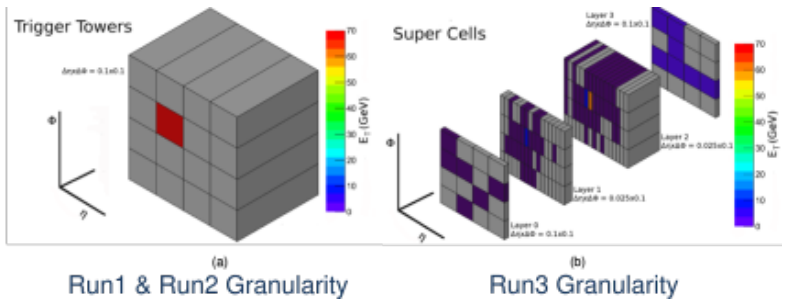


https://cds.cern.ch/record/2719701/files/ATL-PHYS-SLIDE-2020-154.pdf

Upgrades of the LAr Calorimeter System

Phase-I

- higher trigger tower granularity
- new front-end and back-end boards
- reduced trigger rate due to background rejection
- higher geometrical resolution $\longrightarrow$ Lower thresholds and better turn-on curves

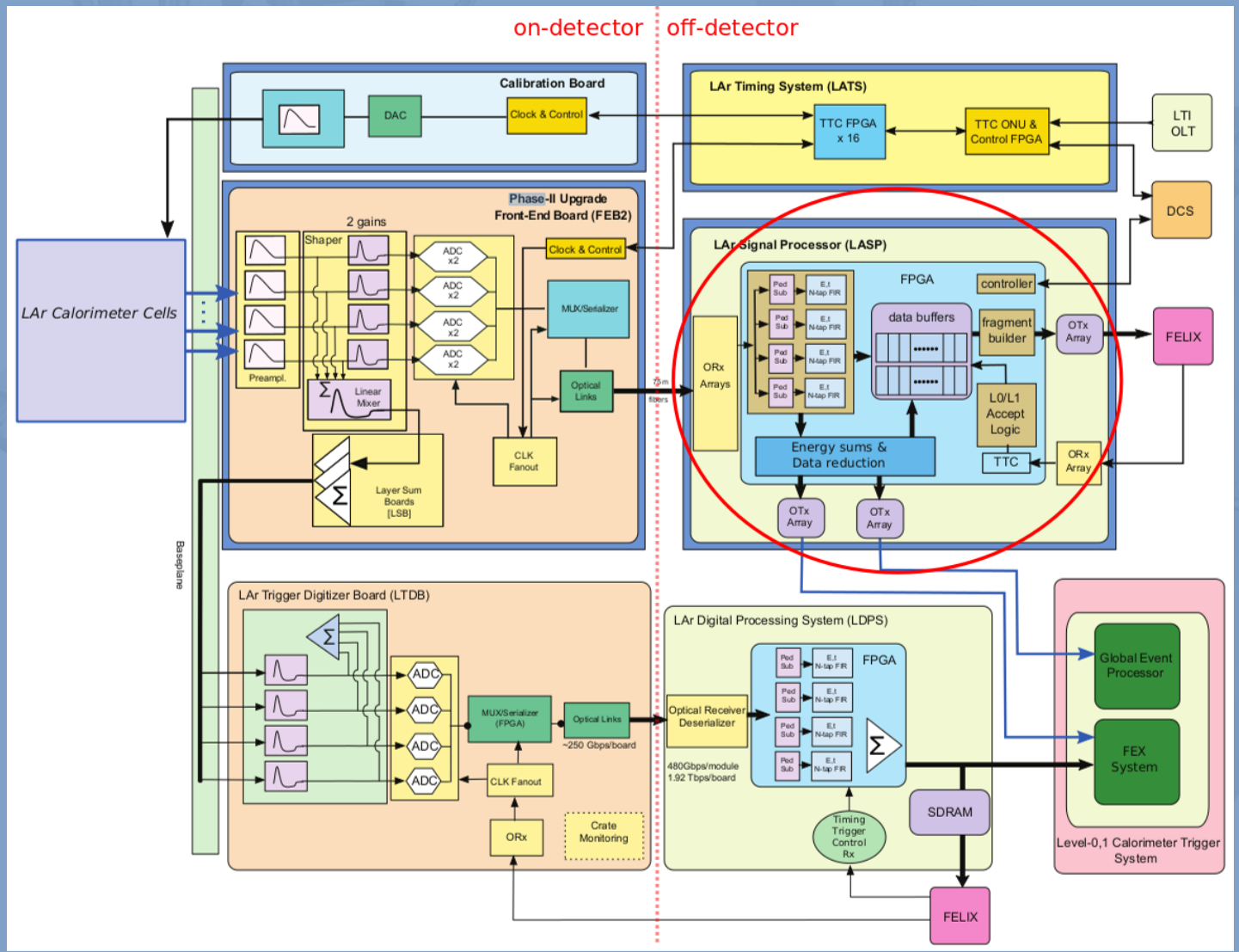


Phase-II

- new front- and back-end electronics to fulfill requirements on latency, trigger rate and radiation hardness
- digitized calorimeter signals will be sent to off-detector electronics with full granularity at 40 MHz

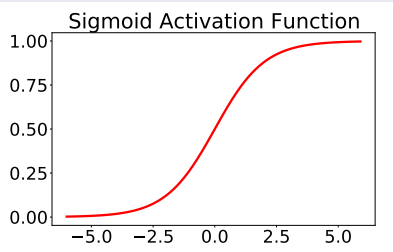
https://cds.cern.ch/record/2719701

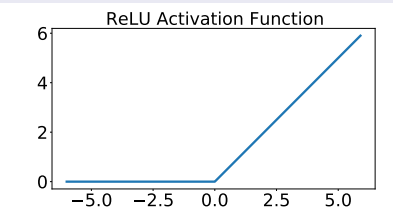
Phase-II Readout Electronics Upgrades

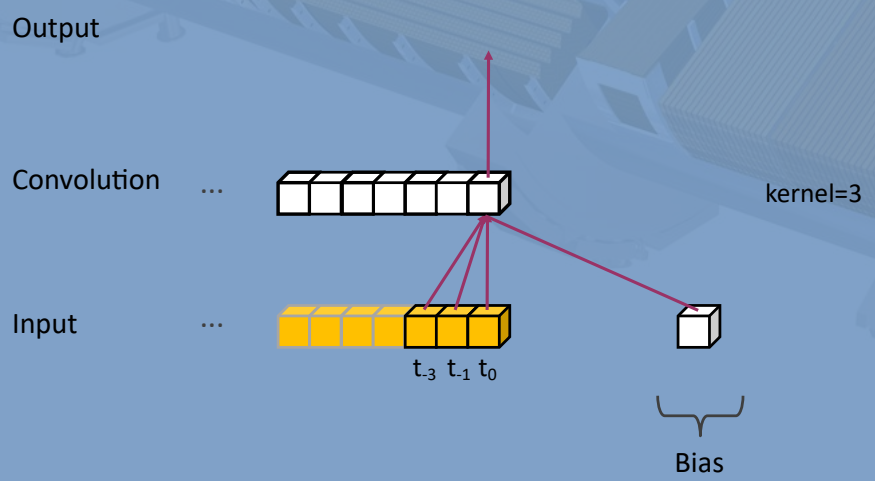


Convolutional Neural Networks (CNNs)

- convolutional operation with certain **filter/kernel** size
- **activation function** gives opportunity to classify, weight, cut

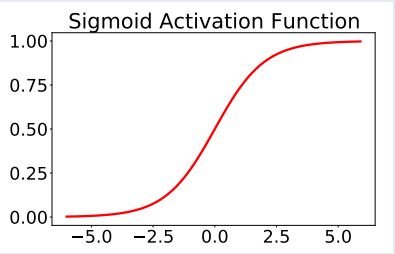


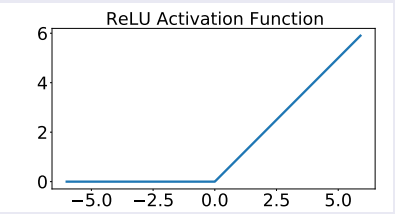


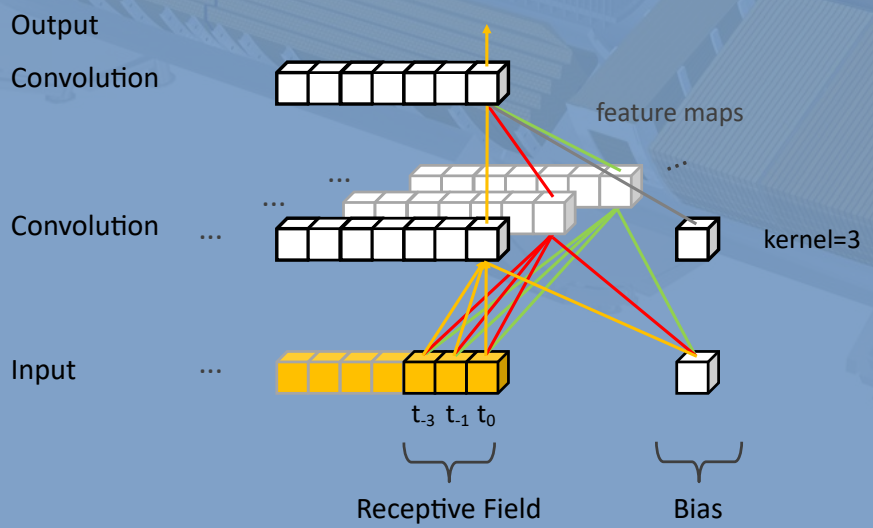


Convolutional Neural Networks (CNNs)

- convolutional operation with certain **filter/kernel** size
- **activation function** gives opportunity to classify, weight, cut
- **feature maps** with different kernels can focus on different properties
- **training** minimizes difference between output and target
- keep number of parameters at minimum ($\sim$ 50-100) since FPGA implementation planned





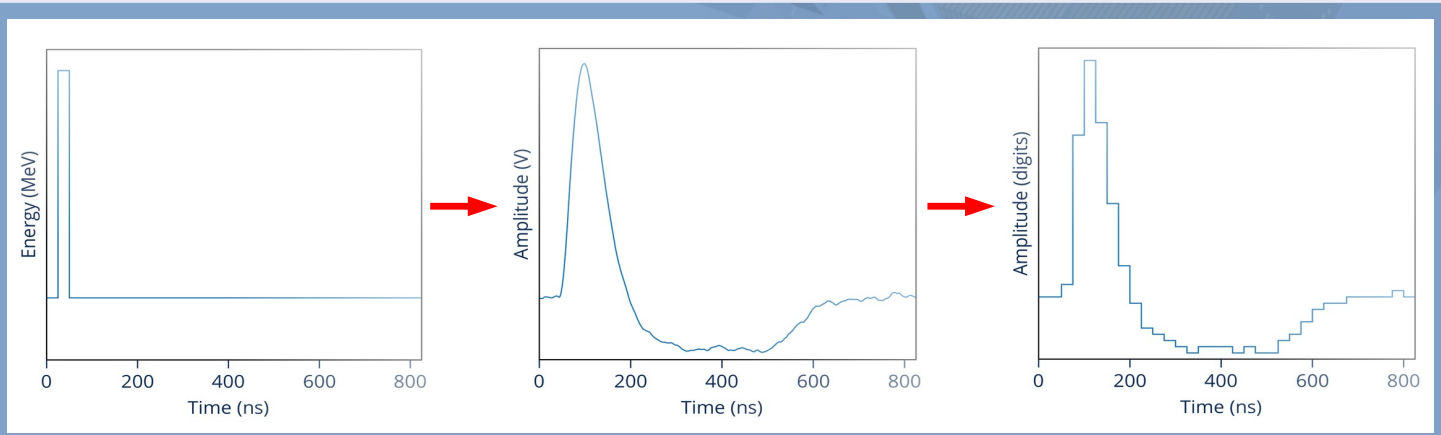


Training Data

AREUS

- AREUS = ATLAS Readout Electronics Upgrade Simulation
- emulates readout chain at LAr calorimeter:
  - event sampling
  - shaping
  - digitization
  - filtering (several filters provided)
- takes analog and digital electronics noise into account
- allows simulation of bunch train patterns

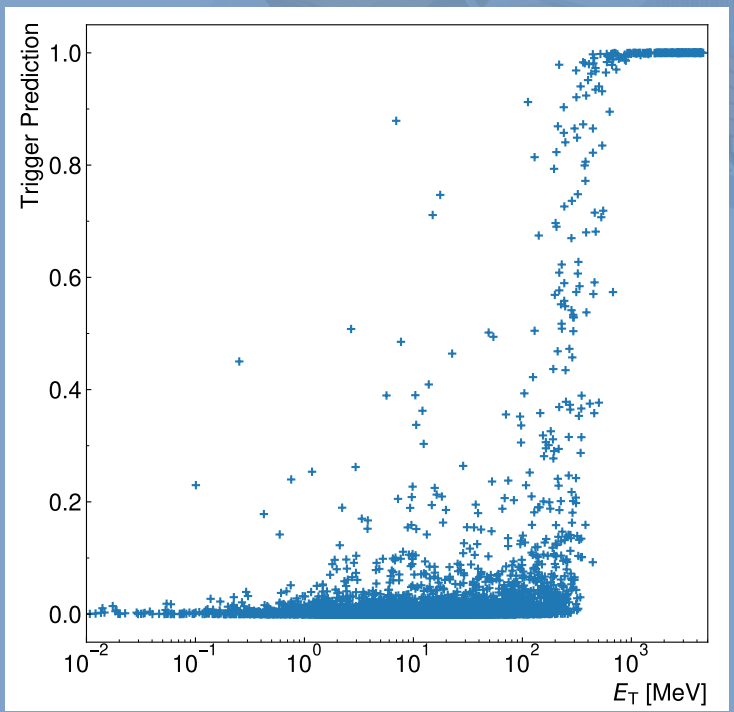
https://gitlab.cern.ch/AREUS

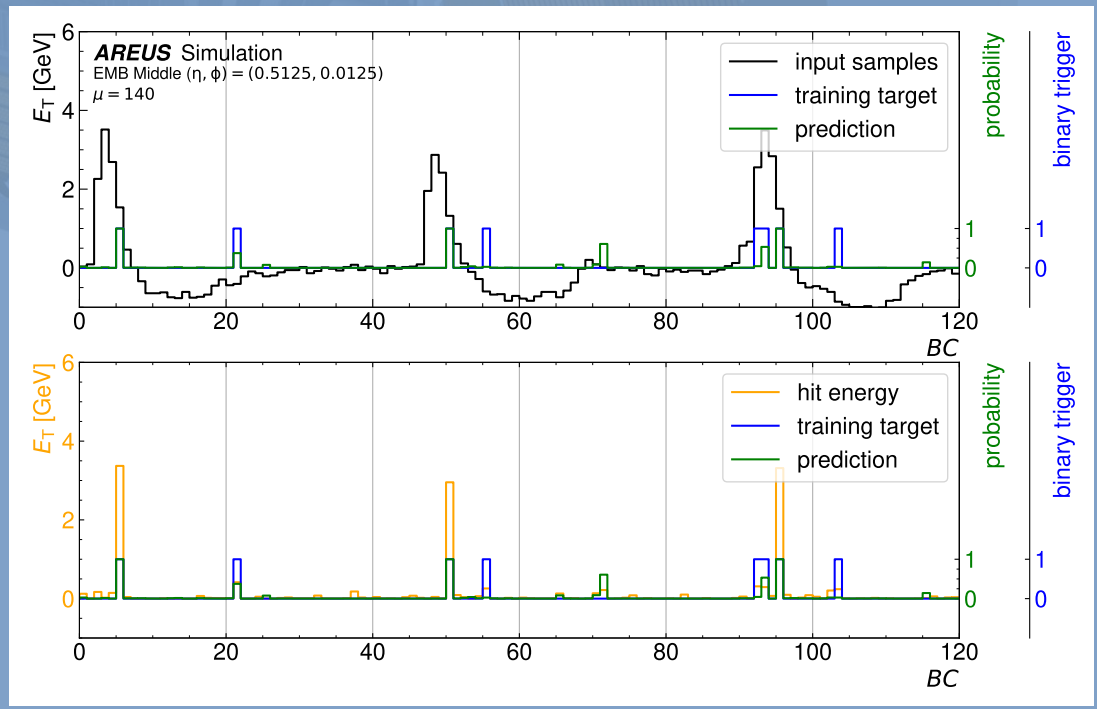


Triggering by **Convolutional Neural Networks**

Neural Net Trigger

- detect *signal hits* from digitized samples
- training on trigger information T:
  - depends on truth energy and threshold E_{th} (about 240 MeV in the LAr barrel region)
  - $T = 1 \iff E_{\mathrm{hit}} > E_{\mathrm{th}}$
  - $T = 0 \iff E_{\mathrm{hit}} \leq E_{\mathrm{th}}$
  - $E_{\mathrm{th}} = 3\sigma$: σ is RMS of electronic noise of particular cell





Energy Reconstruction by **Convolutional Neural Networks**

Our Approach

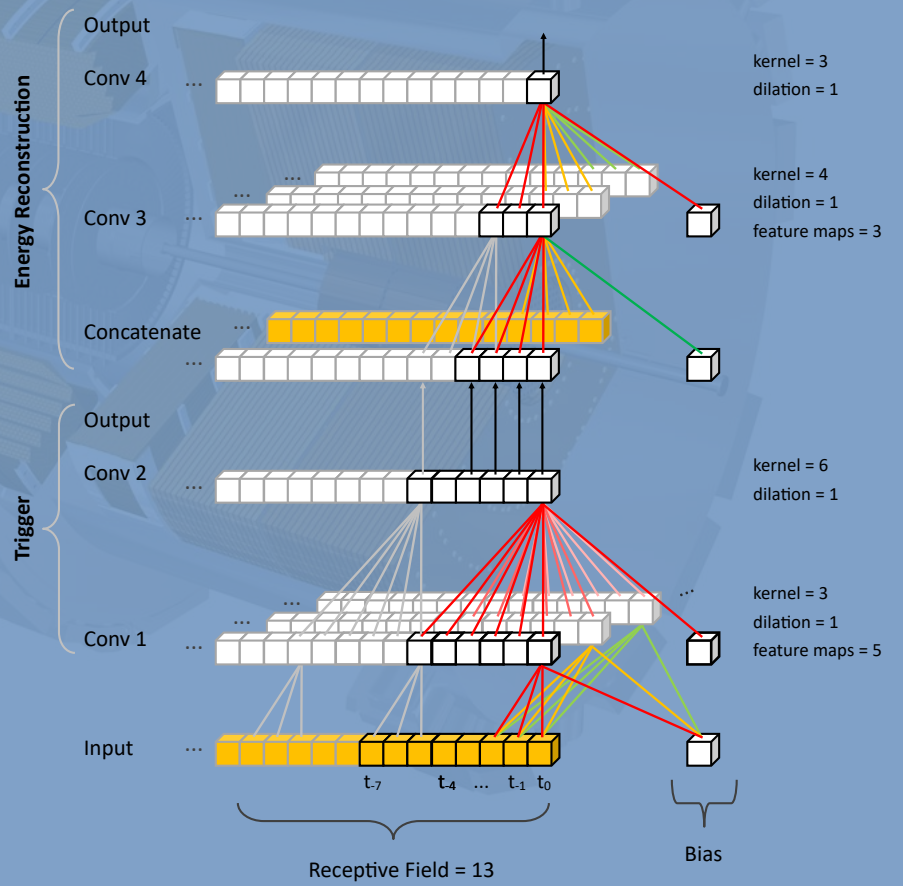
Combine trigger CNN with energy reconstruction CNN

Trigger CNN

- output: probability of detection
- sigmoid activation function
- pretrained parameters

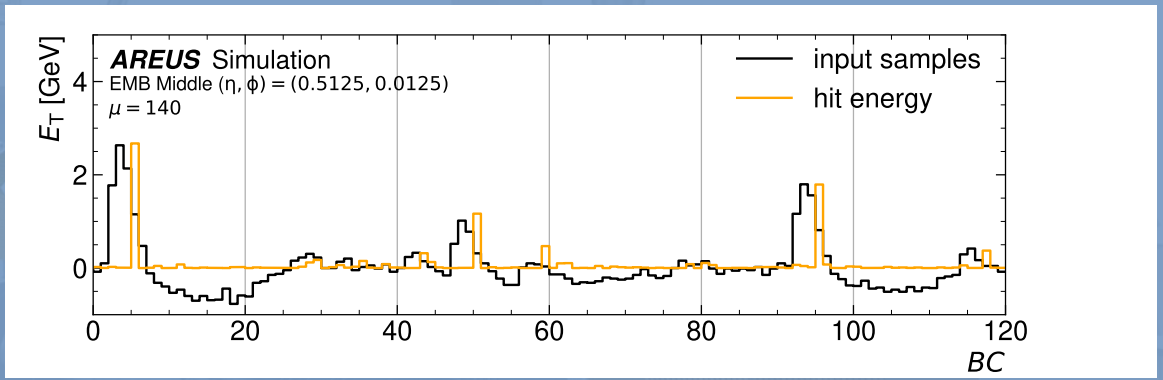
Energy reconstruction CNN

- output: energy
- ReLU activation function
- uses trigger probability and ADC sequence for reconstruction



Training of Convolutional Neural Networks

Input: digitized sequence
Overall target: hit energy



Trigger part:
output is detection probability

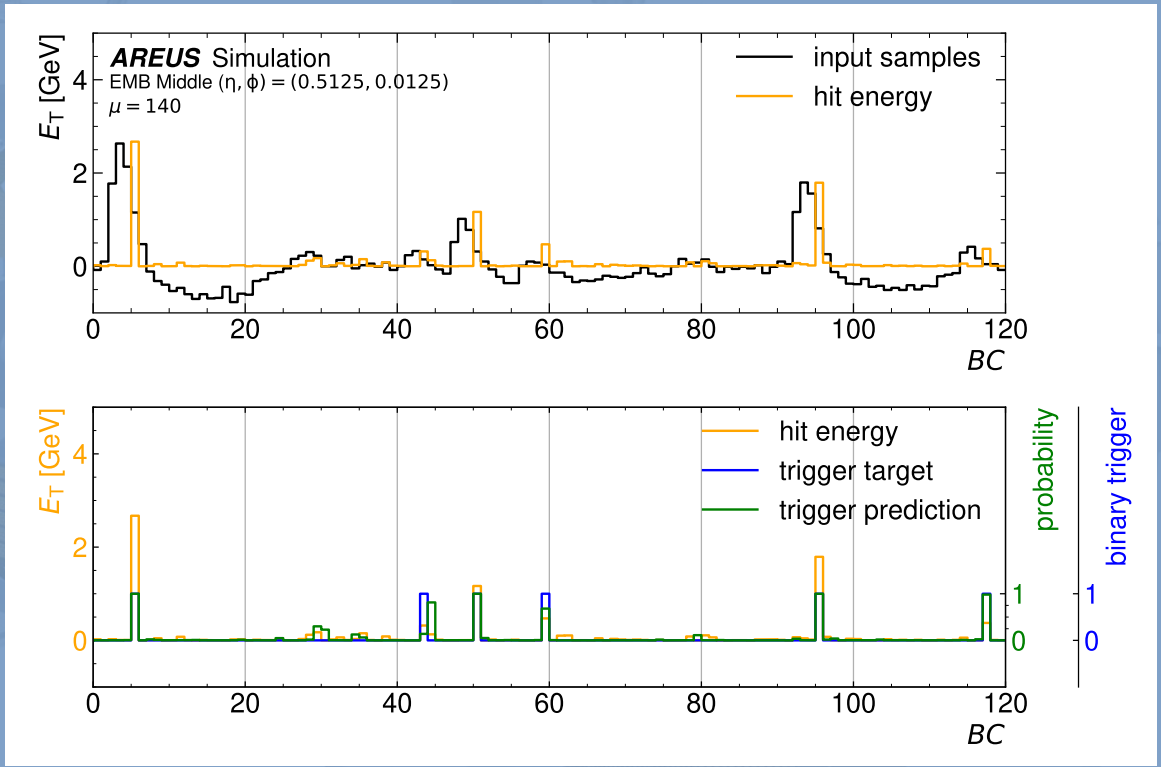
Energy reconstruction part

Training of Convolutional Neural Networks

Input: digitized sequence
Overall target: hit energy

Trigger part:
output is detection probability

Energy reconstruction part

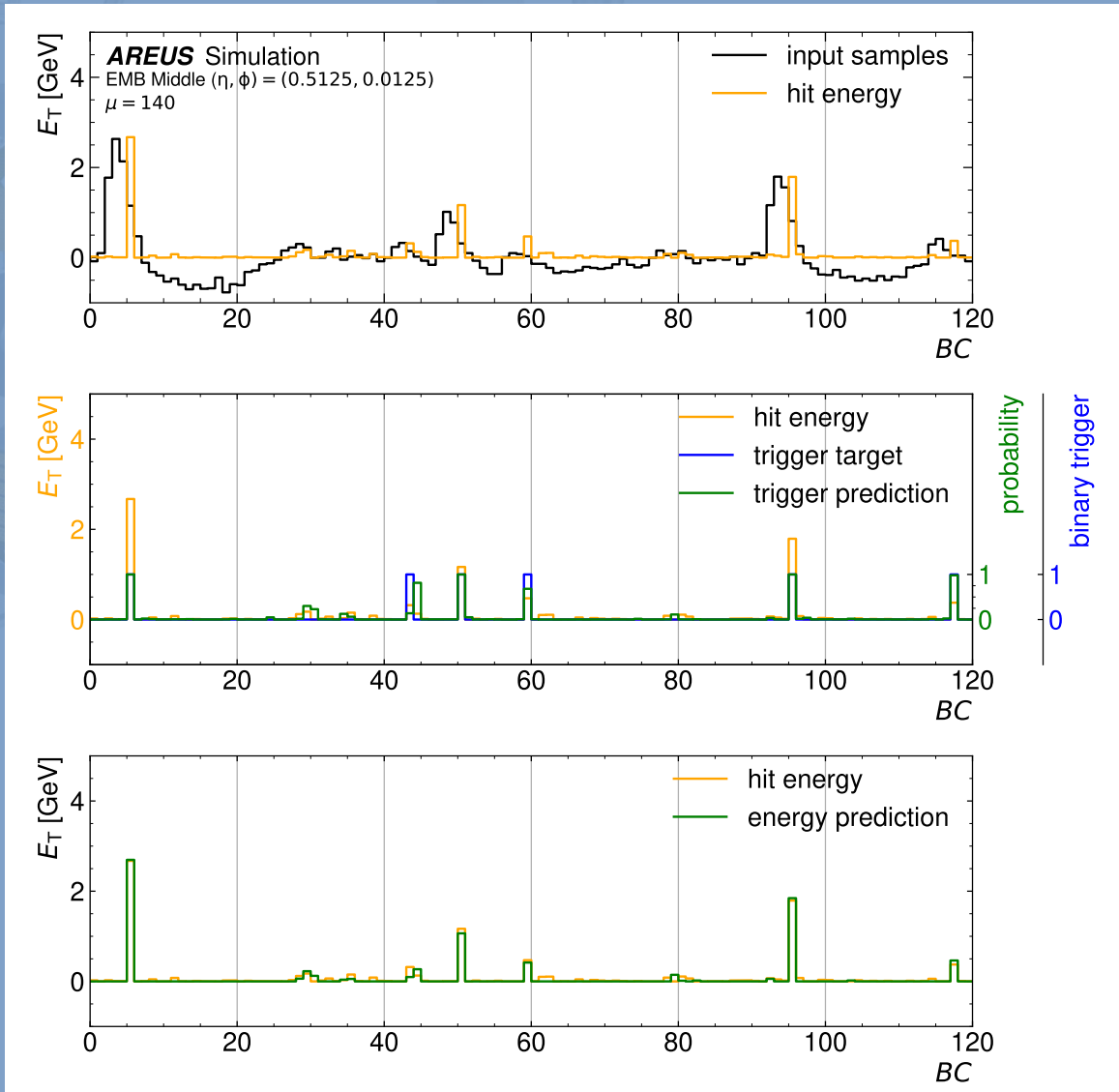


Training of Convolutional Neural Networks

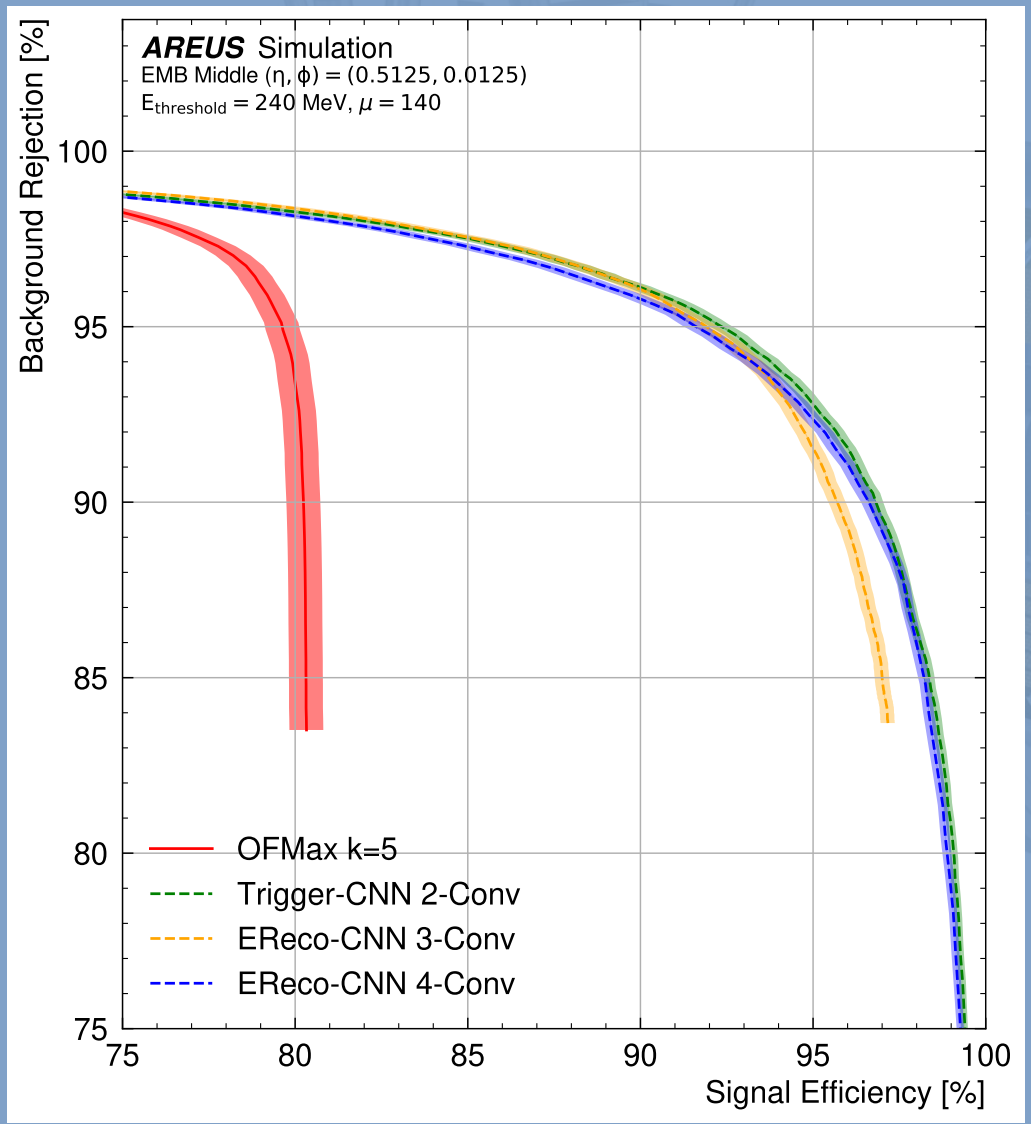
Input: digitized sequence
Overall target: hit energy

Trigger part:
output is detection probability

Energy reconstruction part



Performance Evaluation: OF vs CNN - Trigger Efficiency



ROC curves

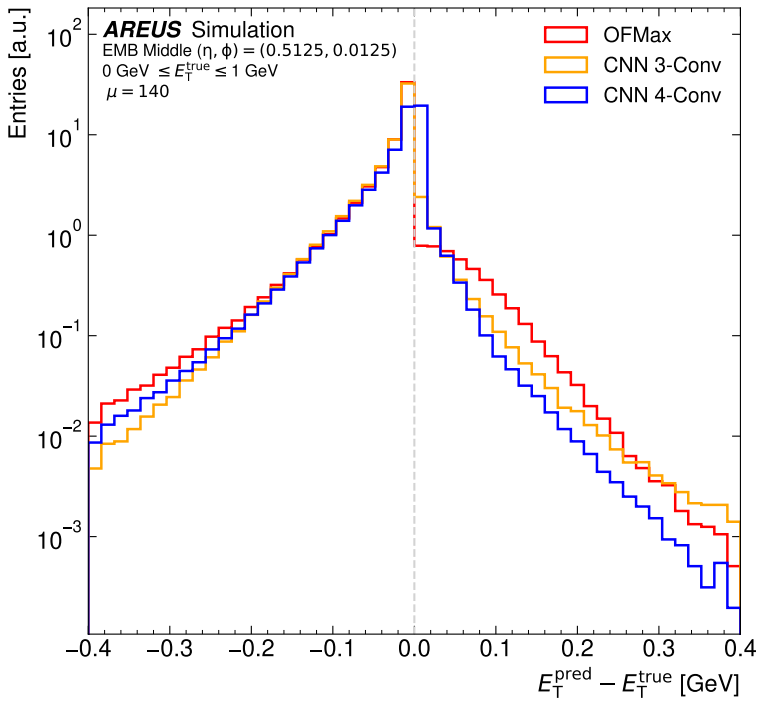
- indicate detection performance
- signal efficiency $= \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$
- background rejection $= \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$
- dependent on threshold

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

Performance Evaluation: OF vs CNN - Energy Resolution

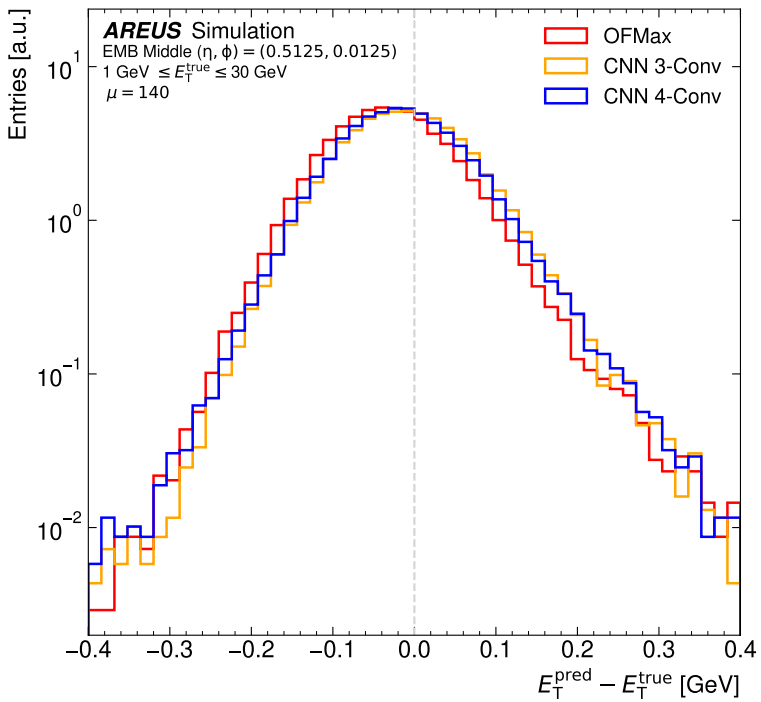
Pile-up Energy Region

0 GeV - 1 GeV



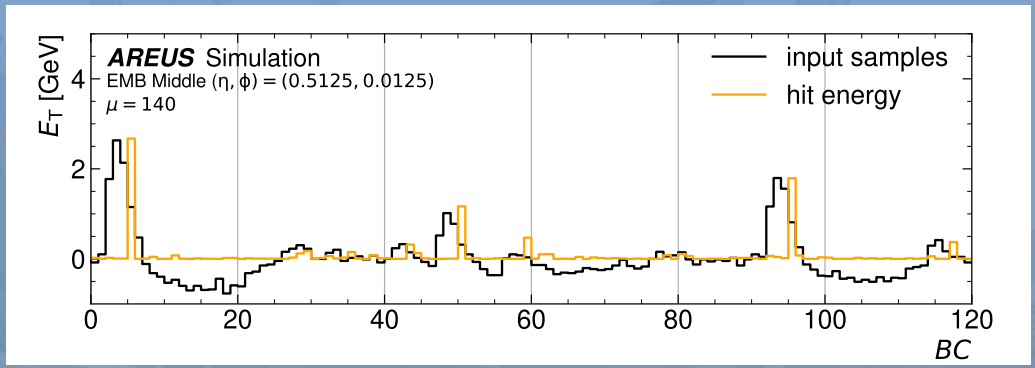
Signal Energy Region

1 GeV - 30 GeV

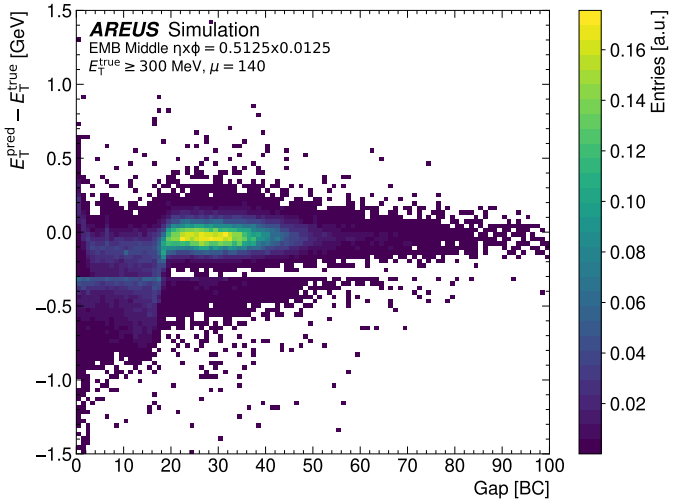


- CNNs improve energy resolution
- CNNs show smaller energy bias especially in low energy region

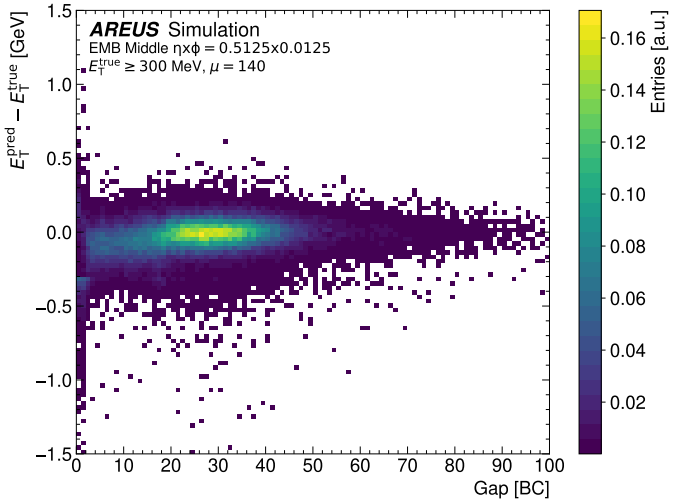
Performance Evaluation: OF vs CNN - Performance on Subsequent Hits



Optimal Filter with Maximum Finder



CNN with 3 Convolutional layers



- CNNs have ability to correctly predict subsequent deposited energies with overlapping pulses

Energy Reconstruction by **Convolutional Neural Networks**

Benchmarks for CNN

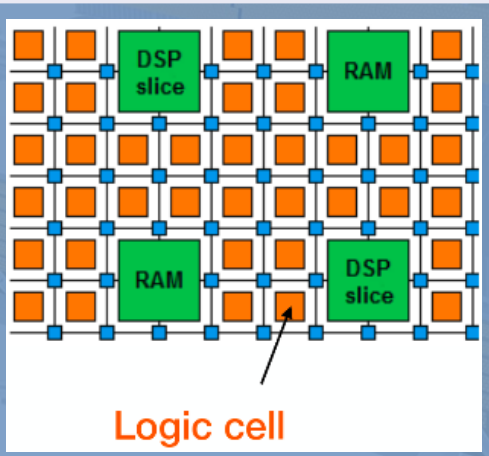
- trigger efficiency must be at an optimum ✓
- energy resolution must at least be as good as with optimal filter ✓
- subsequent hits should be resolvable ✓
- must deal with LHC bunch train structure, gain selection

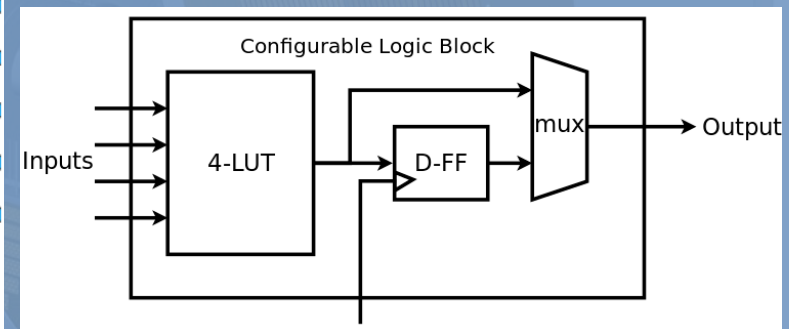
- architecture should be designed such that parameters are kept at a minimum ($\approx$ 50-100) ✓
- FPGA implementation should be possible ...

Digital Signal Processing on Field Programmable Gate Arrays

Field Programmable Gate Arrays (FPGAs)

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





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

Advantages of CNN Implementation on FPGAs

- real time data processing with high frequencies (100 MHz – 1 GHz)
- parallel data processing
- reconfigurable for different CNN structures

FPGA resources for CNN implementation

Digital Signal Processors (DSPs)

- optimized for highspeed multiplications with high bit width
- DSPs may be chained up to perform Multiply-Accumulate operation

$\longrightarrow$ use DSPs for calculating products of layer inputs and weights

Lookup Tables (LUTs)

- configurable for complex logic functions

$\longrightarrow$ use LUTs for non-linear activation functions like sigmoid

Random-Access Memory (RAM)

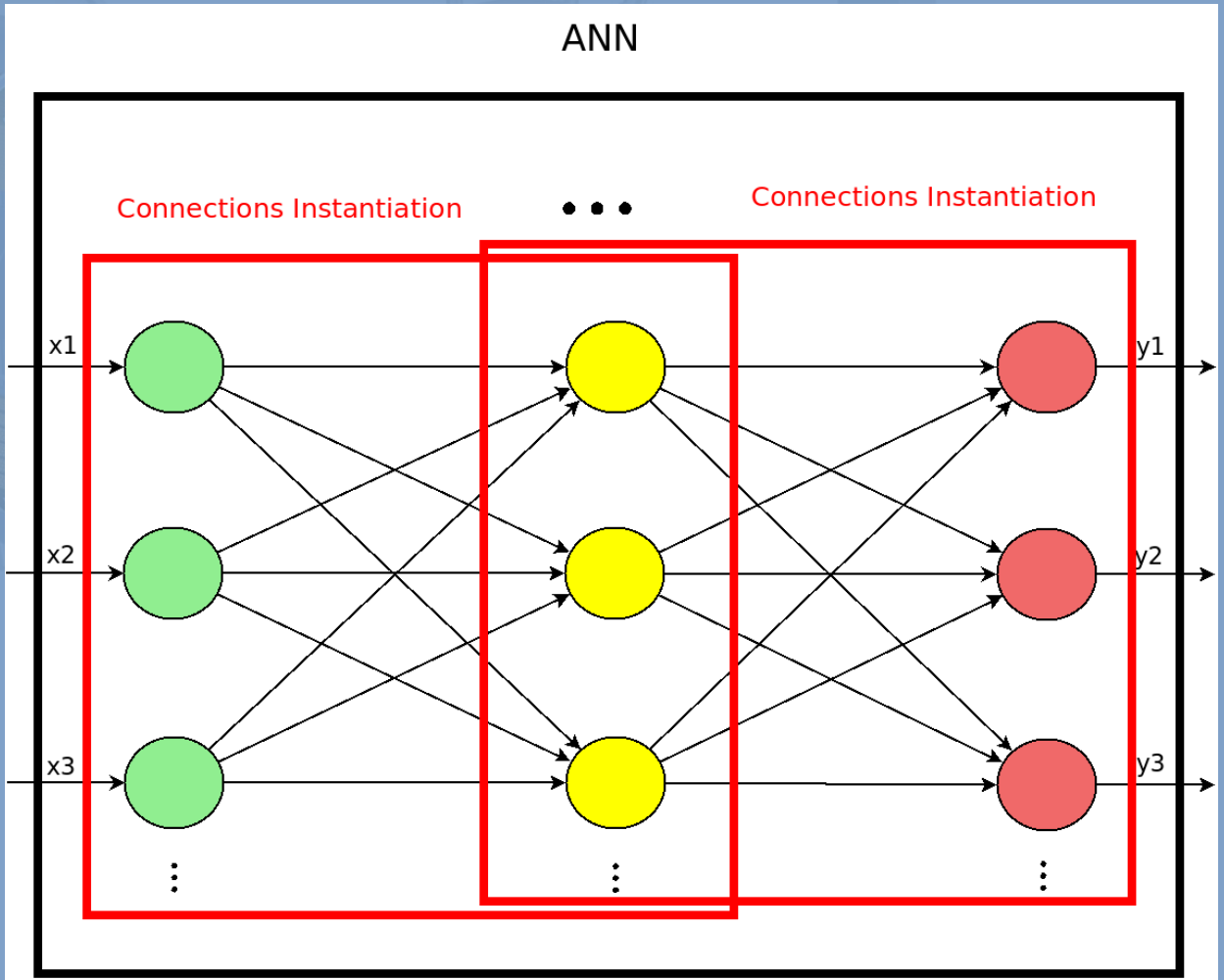
- configurable as dual port RAM
  - one interface to firmware component
  - one interface for data transmission to/from FPGA

$\longrightarrow$ use RAM with slow control system to load new set of weights for the CNN on the run

CNN Firmware Implementation

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)



Pipelining

Firmware Optimization

- aim core frequency of LASP FPGA used for Phase-2 data processing of $480\,\text{MHz} = 12 \times 40\,\text{MHz}$
- minimize latency ($< 150\,\text{ns}$) as trigger accept must come in time
- meet resource limitations of FPGA

$\longrightarrow$ set pipeline registers as a compromise of the factors above

Pipelining over Input Samples

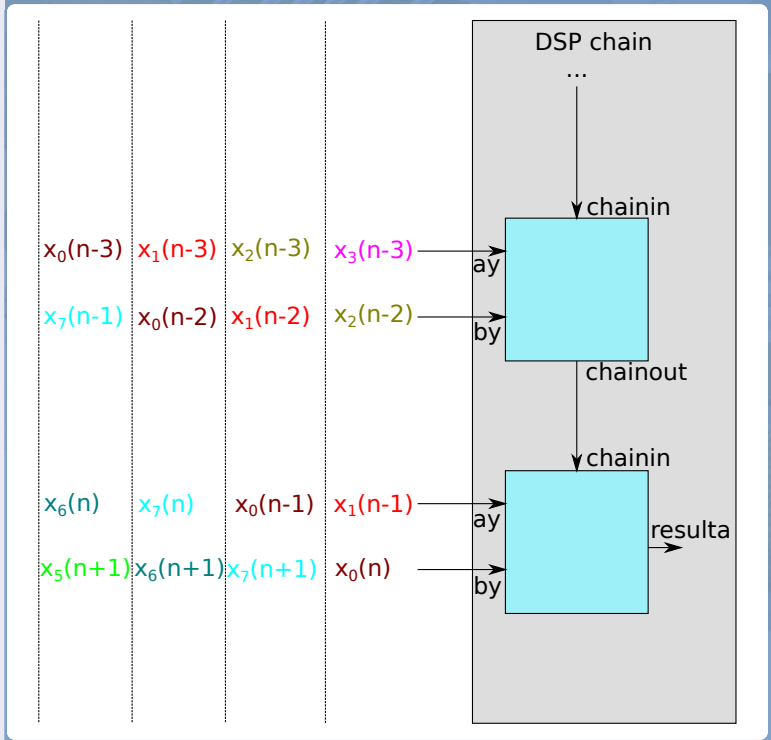
start calculating layer output as soon as first required sample is available

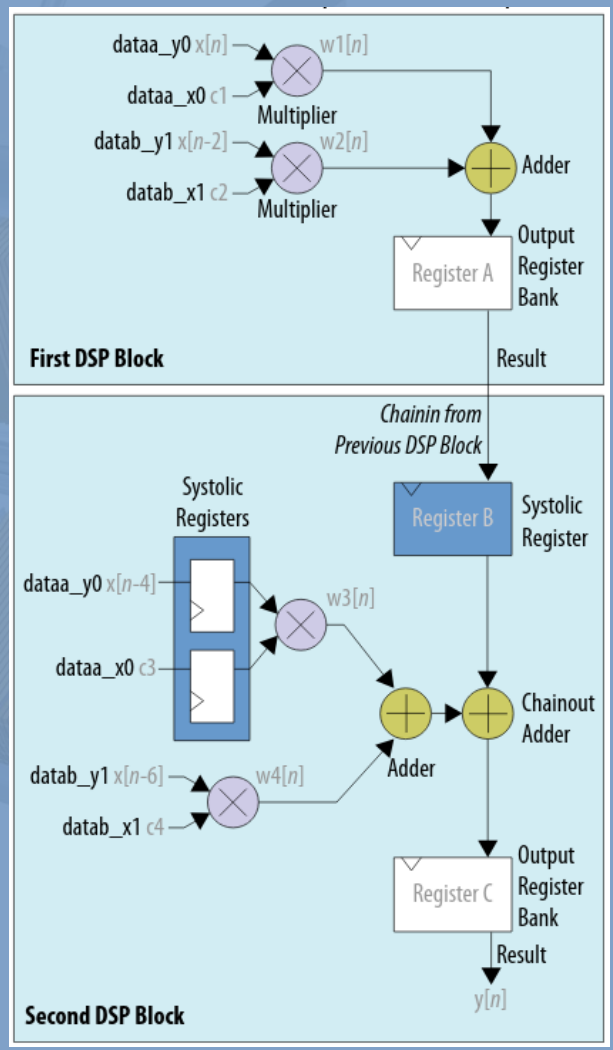
BC	Input			
n	x_0	$y_0 = w_0 \cdot x_0 +$	$w_1 \cdot x_0 +$	$w_2 \cdot x_0 + b$
$n-1$	x_1	$y_1 = w_0 \cdot x_1 +$	$w_1 \cdot x_1 +$	$w_2 \cdot x_1 + b$
$n-2$	x_2	$y_2 = w_0 \cdot x_2 +$	$w_1 \cdot x_2 +$	$w_2 \cdot x_2 + b$

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

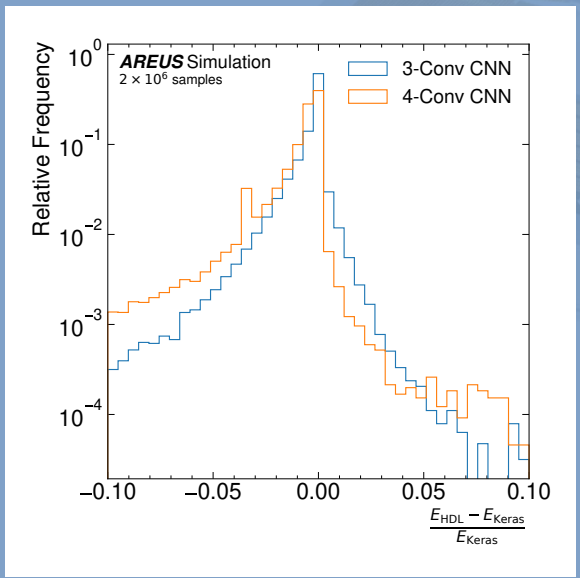


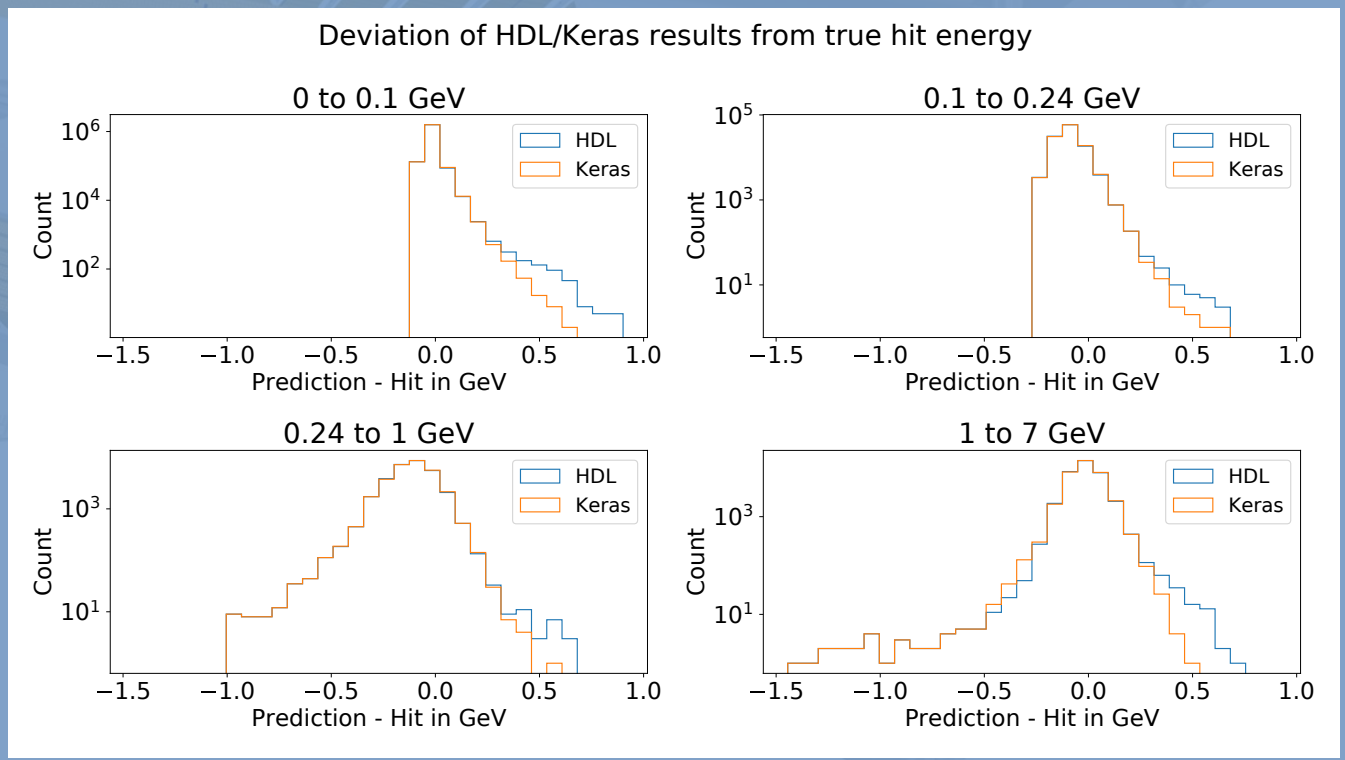


https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/stratix-10/ug-s10-dsp.pdf

Performance

good agreement with floating point software simulation results and true hit energy





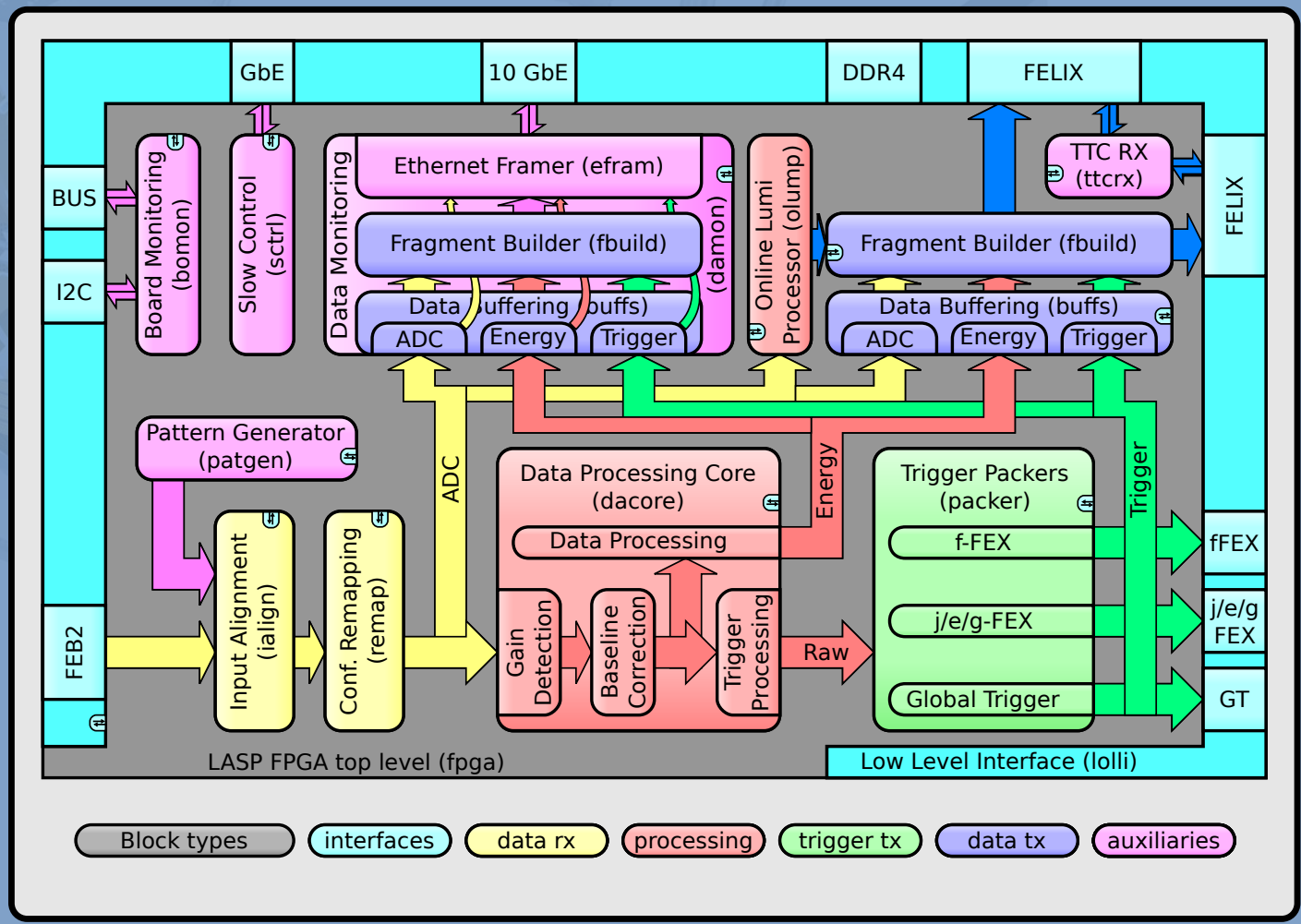
Performance

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

- training and hardware performance must be weighed against each other
- further pipeline optimizations are ongoing

Integration in LAr Signal Processor (LASP)

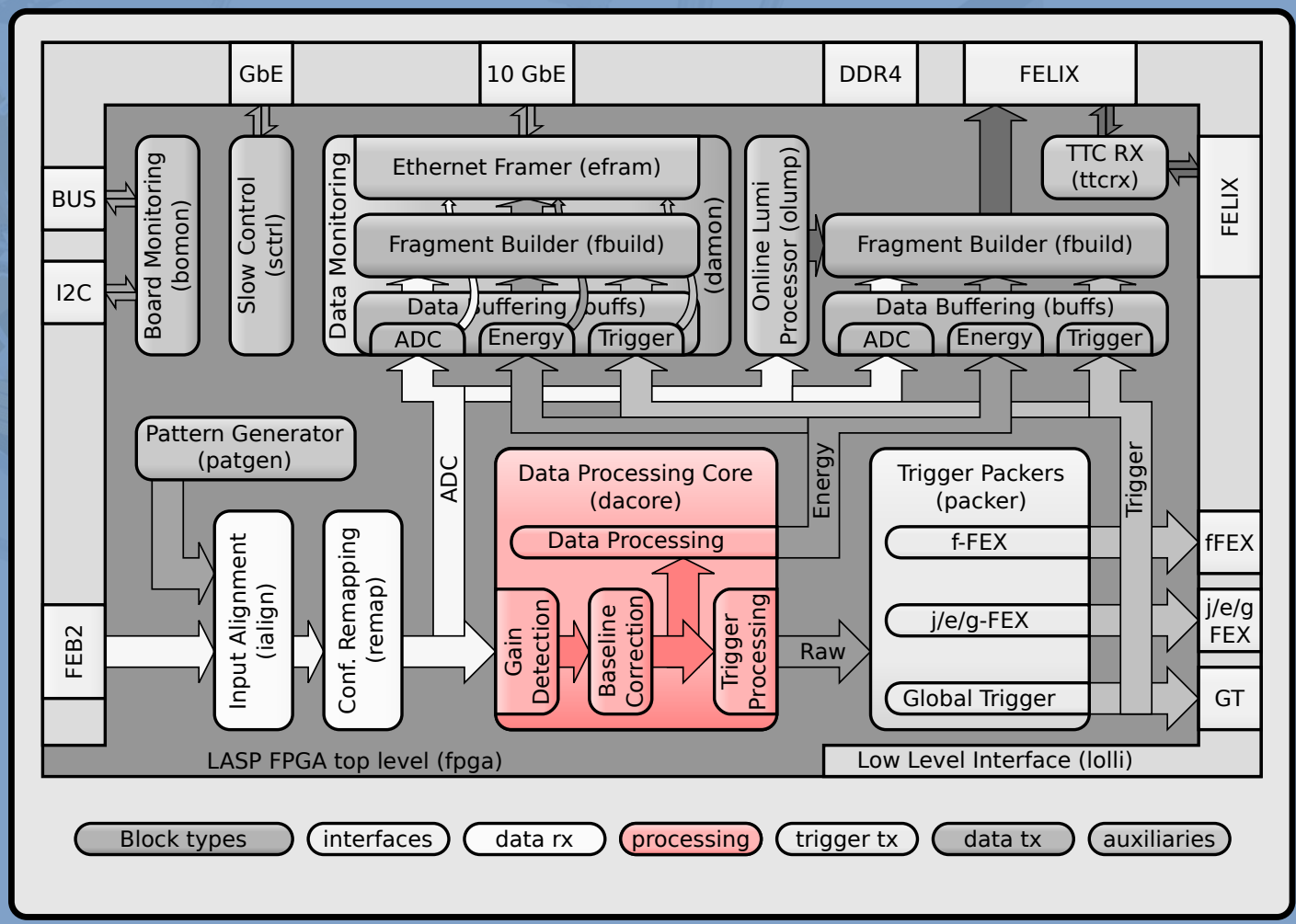
- CNNs to be integrated in data processing core
- outputs data on trigger path, data path and for monitoring



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

Integration in LAr Signal Processor (LASP)

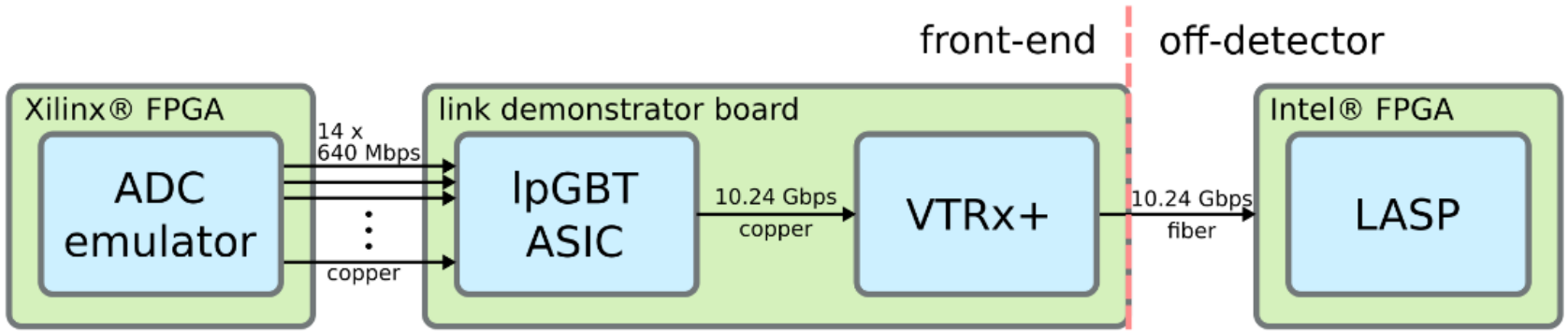
- CNNs to be integrated in data processing core
- outputs data on trigger path, data path and for monitoring

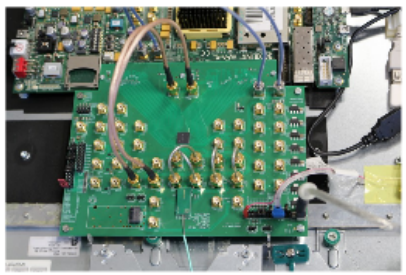


https://gitlab.cern.ch/atlas-lar-be-firmware/LASP/LASP-doc/

Demonstrator for Phase-II Readout Chain

- demonstrator of Phase-II readout chain in Dresden is developed by Markus Helbig
- emulates signal from FEB and sends it through link demonstrator board to LAr Signal Processor





[https://www.xilinx.com/content/dam/xilinx/imgs/kits/whats-inside/VC707-evaluation-board.jpg]

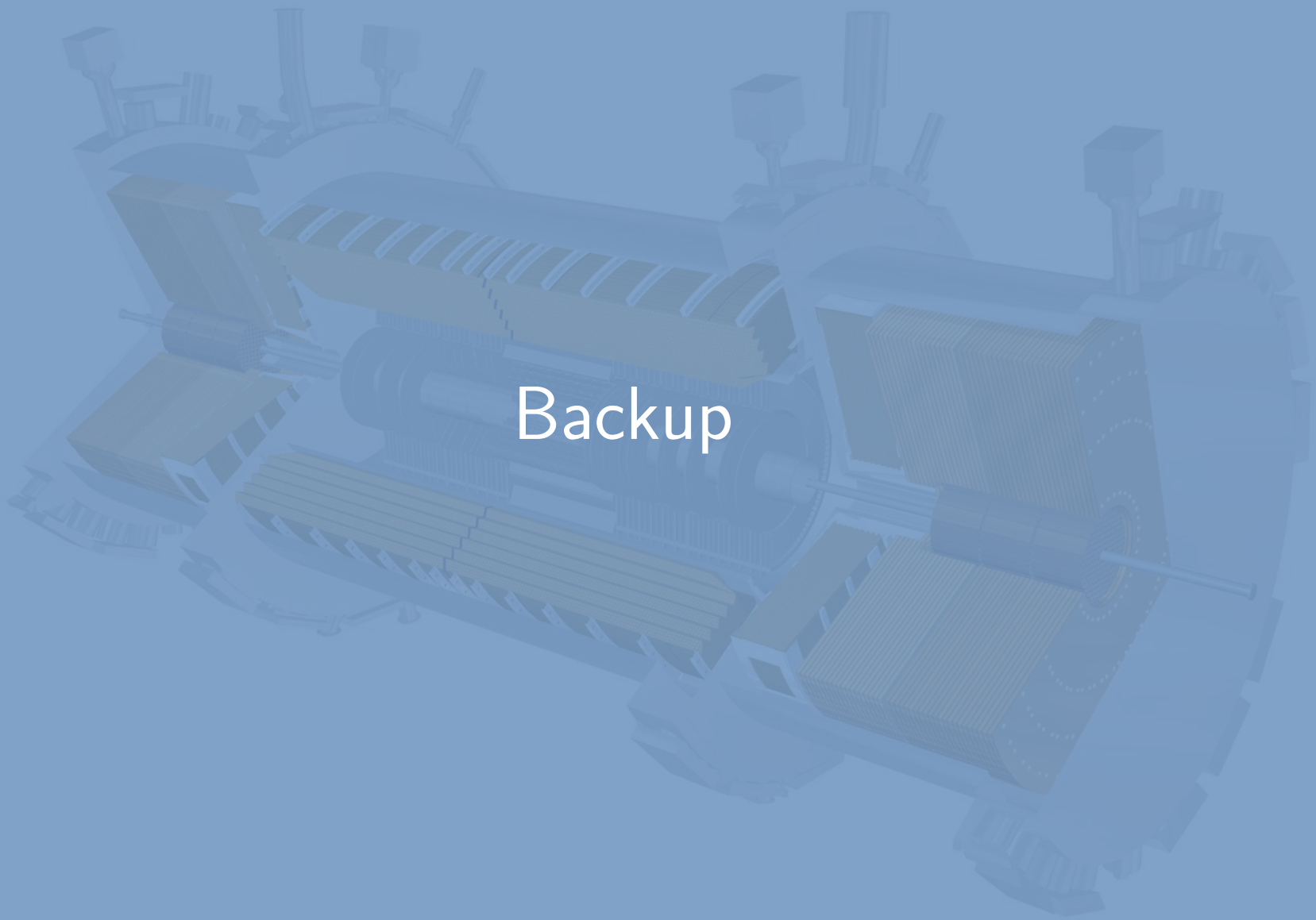
[https://www.mouser.de/images/marketingid/2018/img/153176375_Intel_Stratix-10-Development-Kits.png]

Summary

CNNs for Energy Reconstruction

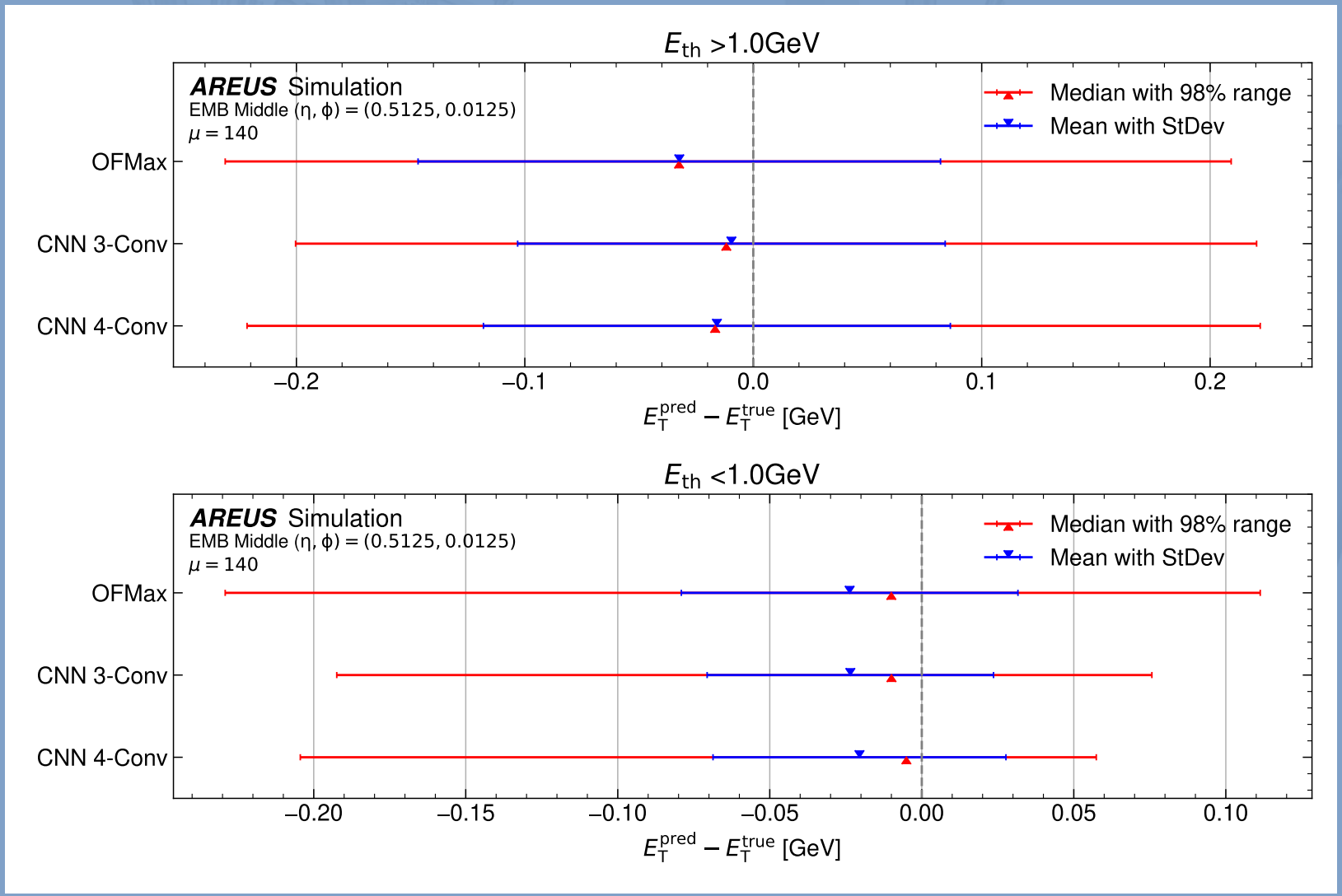
- electronics of ATLAS LAr Calorimeters will be upgraded for HL-LHC until 2027
- harsh environment with up to 200 pile-up events
- energy reconstruction by convolutional neural networks shows promising results
- CNN implementation in FPGA successful
- next steps:
  - study network performance for even more realistic cases: bunch trains, correlated noise events, ...
  - optimize FPGA implementation and run full processing chain on hardware

Thanks for your attention!



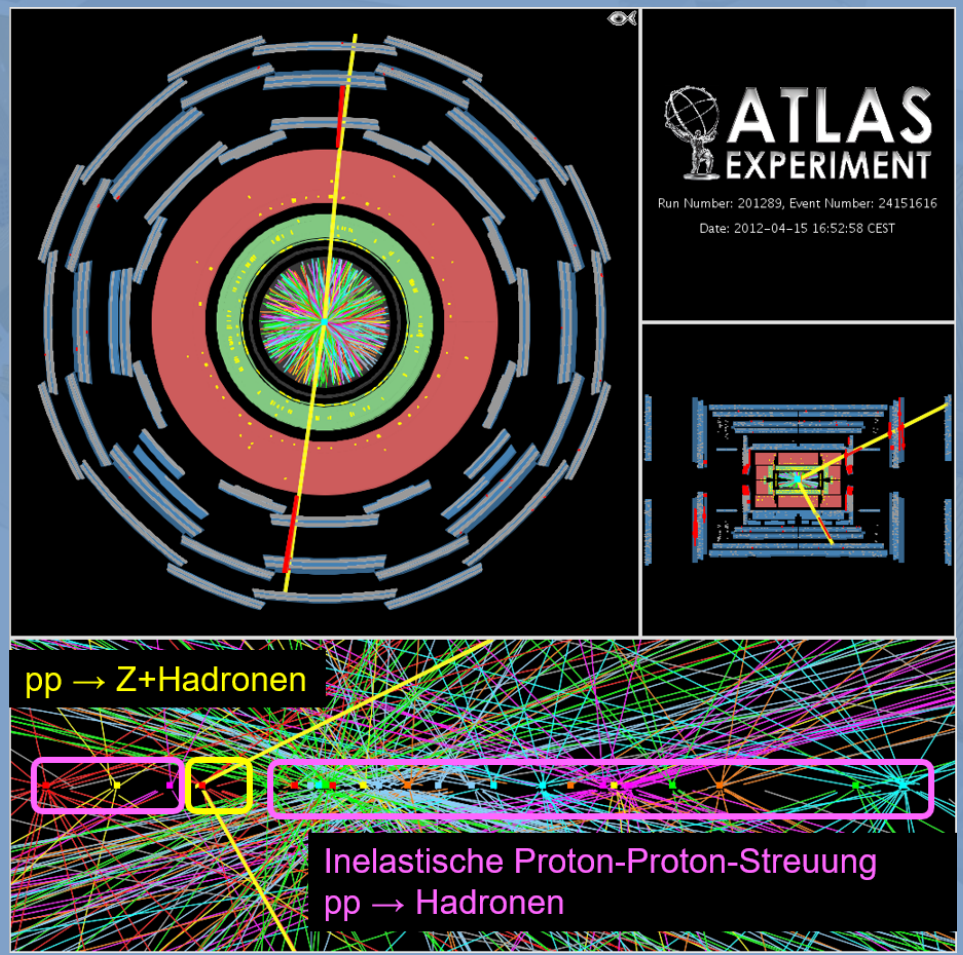
Backup

Performance Evaluation: OF vs CNN - Energy Resolution



Event Display with Pileup

Real Event from 2012



https://iopscience.iop.org/article/10.1088/1742-6596/523/1/012018/pdf