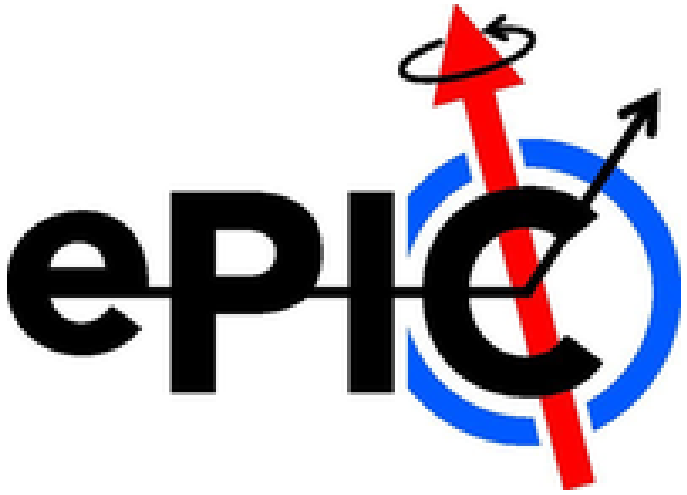
# ePIC Italia
# General Meeting

## Update on data compression and SRO WG activities

Fabio Rossi



# SRO WG Overview

ePIC DAQ uses streaming architecture

## DAQ

Definition of streaming is “No L0 trigger”

- All data is zero suppressed by the front end electronics
- No system wide deadtime in normal operation
- Collaboration should have the full ability to make data selection cuts on the widest possible criteria
  - Flexible event selection, data selection and background characterization
- But subject to an overall throughput budget of ~100Gb/sec

ePIC Streaming will include

- Capabilities for hardware and firmware based triggering
- Capability for flow control
- Zero suppression & aggregation within data packets

Greater sensitivity to noise

Greater sensitivity to backgrounds

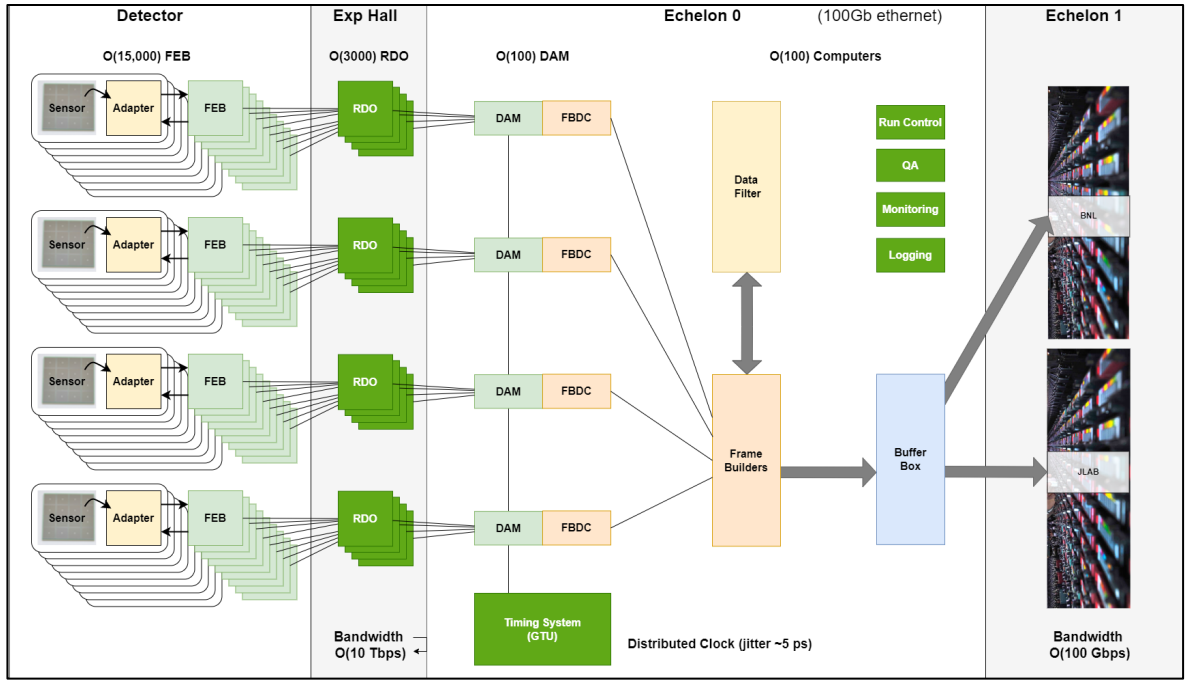
## Computing

Definition of streaming is “Process data as it arrives”

- Fast Analysis (~3 weeks not months or years) using automation of calibration and reconstruction.

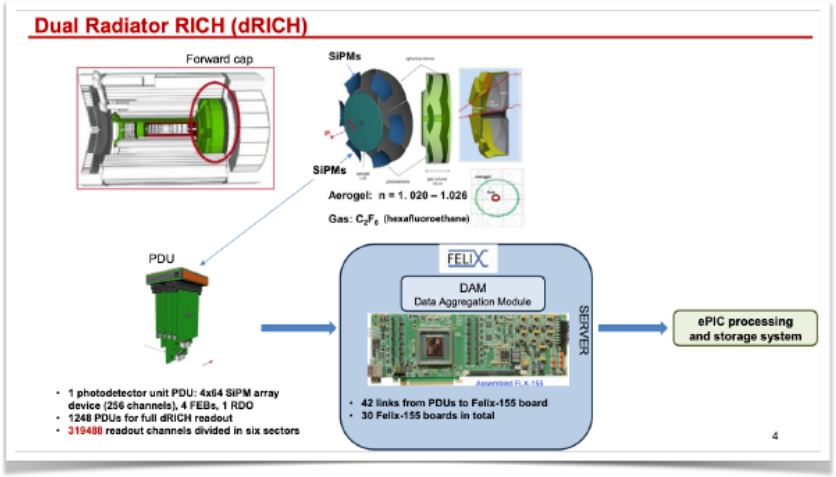
Requires some overlap between DAQ/computing

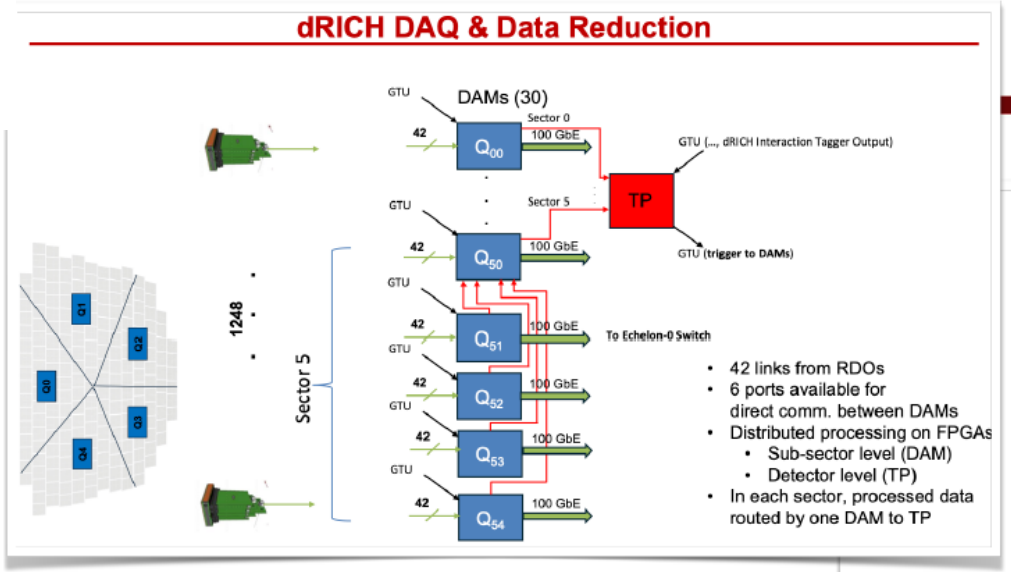
- Automation of calibrations
  - Mapping calibration dependencies
  - Mapping application of calibrations
  - Mapping evolution of calibrations
- QA and monitoring can make use of full offline structure
- Consistent schemes and language for data/metadata
- Event selection / tagging / and accounting

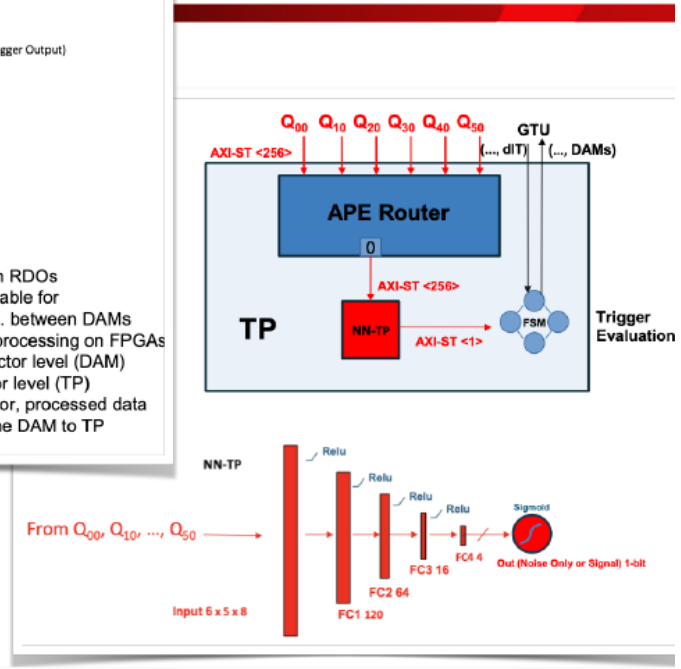


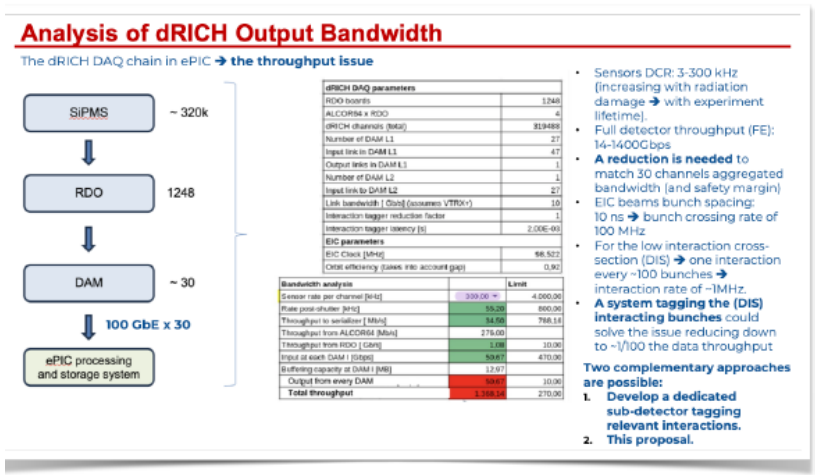

- Organizational contribution
- SRO Prototypes
- Data reduction
- dRICH DAQ

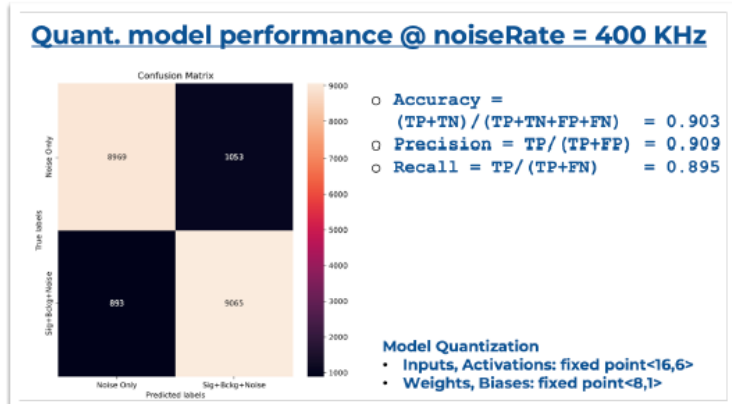
# dRICH

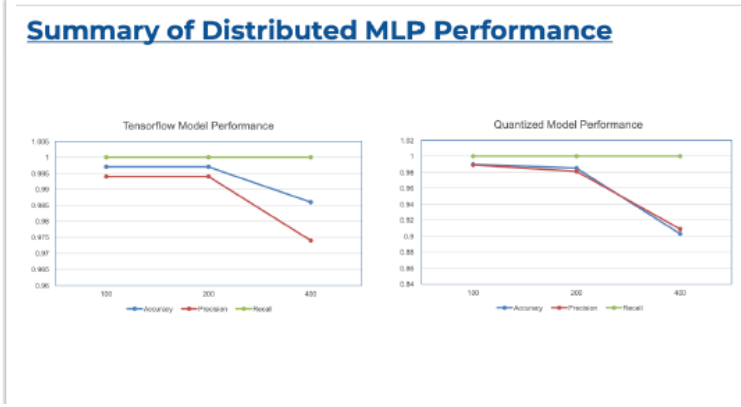

Dual Radiator RICH (dRICH)
Forward cap
SiPMs
SiPMs
Aerogel: n = 1. 020 – 1.026
Gas: $C_2F_6$ (hexafluoroethane)
PDU
FELIX
DAM
Data Aggregation Module
SERVER
ePIC processing and storage system
1 photodetector unit PDU: 4x64 SiPM array device (256 channels), 4 FEBs, 1 RDO
1248 PDUs for full dRICH readout
319488 readout channels divided in six sectors
42 links from PDUs to Felix-155 board
30 Felix-155 boards in total
4


dRICH DAQ & Data Reduction
DAMs (30)
Sector 0
Sector 5
Sector 5
1248
GTU
42
100 GbE
$Q_{00}$
$Q_{50}$
$Q_{51}$
$Q_{52}$
$Q_{53}$
$Q_{54}$
TP
GTU (..., dRICH Interaction Tagger Output)
GTU (trigger to DAMs)
To Echelon-0 Switch
42 links from RDOs
6 ports available for direct comm. between DAMs
Distributed processing on FPGAs
Sub-sector level (DAM)
Detector level (TP)
In each sector, processed data routed by one DAM to TP


$Q_{00}$ $Q_{10}$ $Q_{20}$ $Q_{30}$ $Q_{40}$ $Q_{50}$
GTU
(..., dIT)
(..., DAMs)
AXI-ST <256>
APE Router
0
AXI-ST <256>
TP
NN-TP
AXI-ST <1>
FSM
Trigger Evaluation
NN-TP
From $Q_{00}$, $Q_{10}$, ..., $Q_{50}$
Input 6 x 5 x 8
Relu
Relu
Relu
Relu
Sigmoid
FC1 120
FC2 64
FC3 16
FC4 4
Out (Noise Only or Signal) 1-bit


Analysis of dRICH Output Bandwidth
The dRICH DAQ chain in ePIC → the throughput issue
SiPMS
~ 320k
RDO
1248
DAM
~ 30
100 GbE x 30
ePIC processing and storage system
Sensors DCR: 3-300 kHz (increasing with radiation damage → with experiment lifetime).
Full detector throughput (FE): 14-1400Gbps
A reduction is needed to match 30 channels aggregated bandwidth (and safety margin)
EIC beams bunch spacing: 10 ns → bunch crossing rate of 100 MHz
For the low interaction cross-section (DIS) → one interaction every ~100 bunches → interaction rate of ~1MHz.
A system tagging the (DIS) interacting bunches could solve the issue reducing down to ~1/100 the data throughput
Two complementary approaches are possible:
1. Develop a dedicated sub-detector tagging relevant interactions.
2. This proposal.

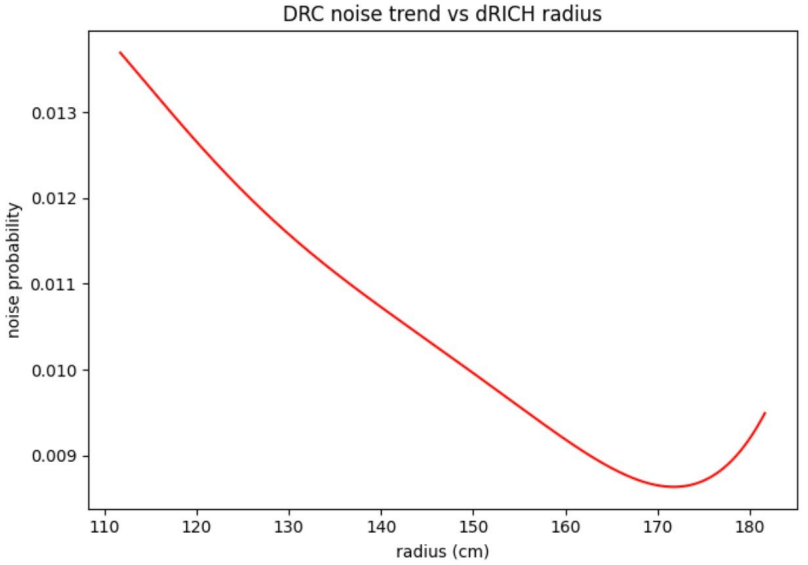

Quant. model performance @ noiseRate = 400 KHz
Confusion Matrix
Noise Only
Sig+Bckg+Noise
True labels
Predicted labels
8969
1053
893
9065
Accuracy = (TP+TN) / (TP+TN+FP+FN) = 0.903
Precision = TP/ (TP+FP) = 0.909
Recall = TP/ (TP+FN) = 0.895
Model Quantization
Inputs, Activations: fixed point<16,6>
Weights, Biases: fixed point<8,1>


Summary of Distributed MLP Performance
Tensorflow Model Performance
Quantized Model Performance
Accuracy
Precision
Recall

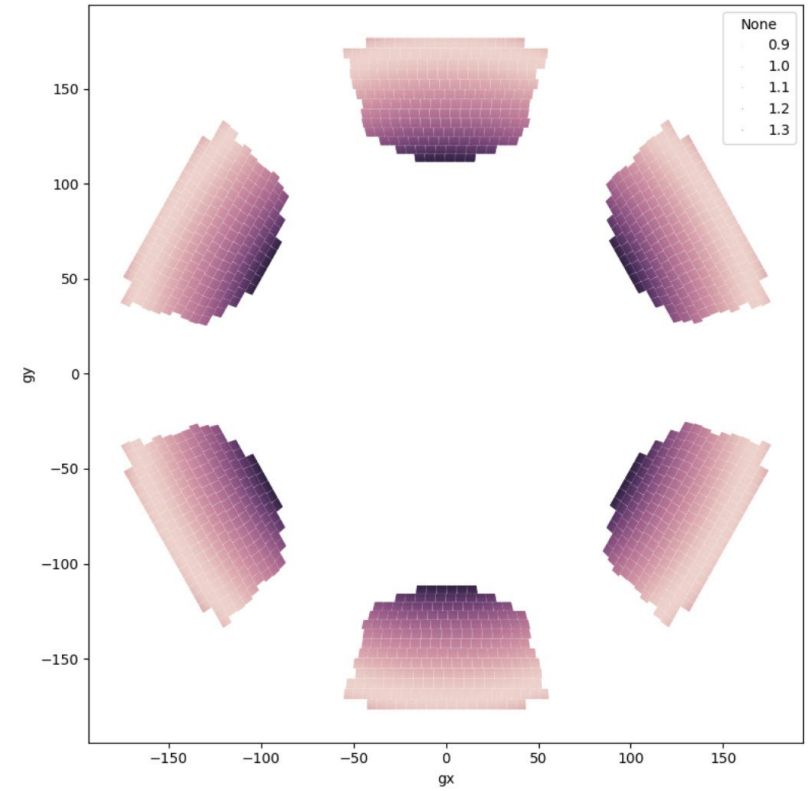
# Realistic Noise Model

Thanks to R. Preghenella's contribution, we added to the toy noise model ("gaussian") a new "realistic" one in which the Dark Count probability of a certain dRICH SiPM is dependent on its radial distance from the detector z-axis and on the integrated luminosity

**⇒ Implemented in EICRecon digitization step (new flag to enable new model noise)**



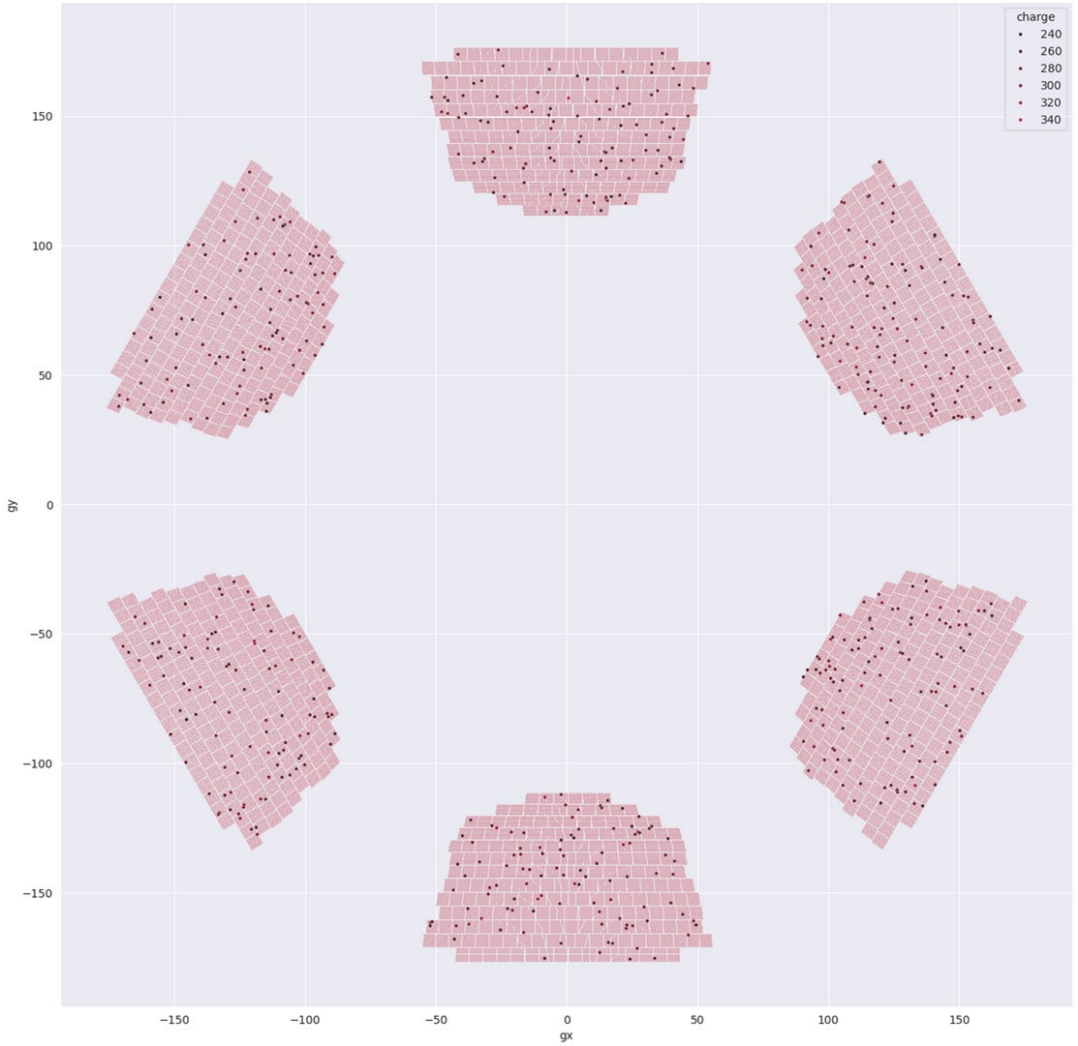

```
const float baseline_dcr = 3.e3; // [Hz] new sensors at T = -30 C and Vover = 4V
const float dcr_increase = 300.e3 / 1.e9; // [Hz/neq]
float neq_radius_params[6] = { -3.27029e+09, 1.26055e+08, -1.88568e+06, 13929.1, -50.9931, 0.0741068 };

float neq_radius(float radius /* cm */)
{
  float neq = 0.;
  for (int ipar = 0; ipar < 6; ++ipar)
    neq += neq_radius_params[ipar] * std::pow(radius, ipar);
  return neq;
}

float
noise_probability(float radius = 150. /* cm */, float window = 10. /* ns */, float luminosity = 100. /* fb-1 */)
{
  float neq = neq_radius(radius) * luminosity;
  float dcr = baseline_dcr + dcr_increase * neq;
  float pro = dcr * window; //* 1.e-9;
  return pro;
}
```




Reconfigurable parameters, used for the new dataset:

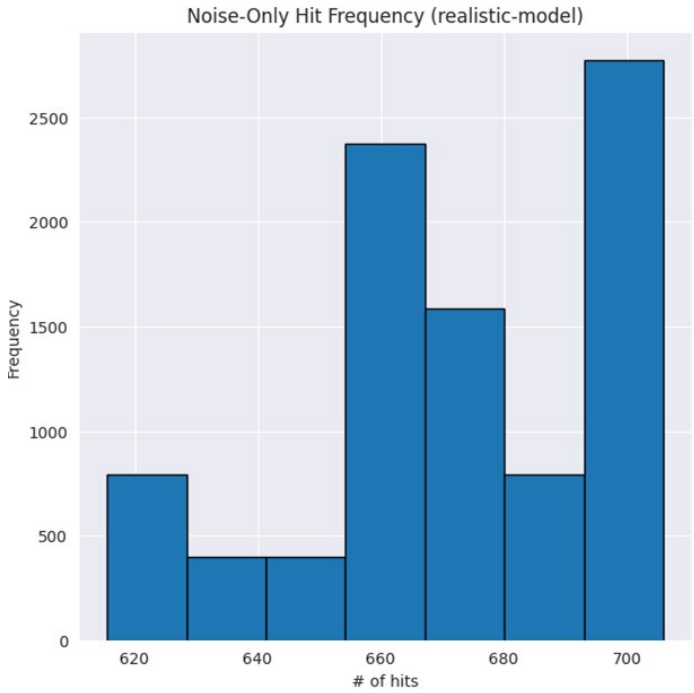
- (Noise) Time window = 10 ns
- (Integrated) Luminosity = 2 fb-1

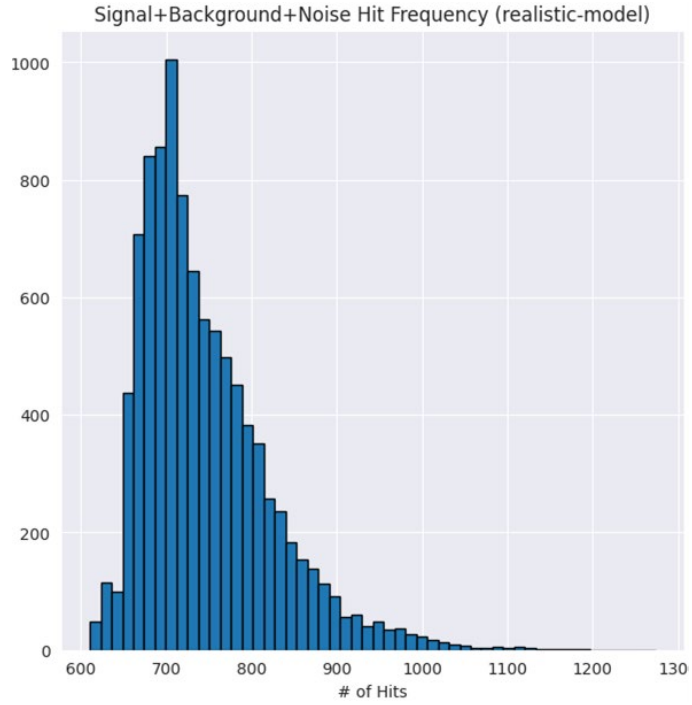
# Updates to the Data Reduction Pipeline



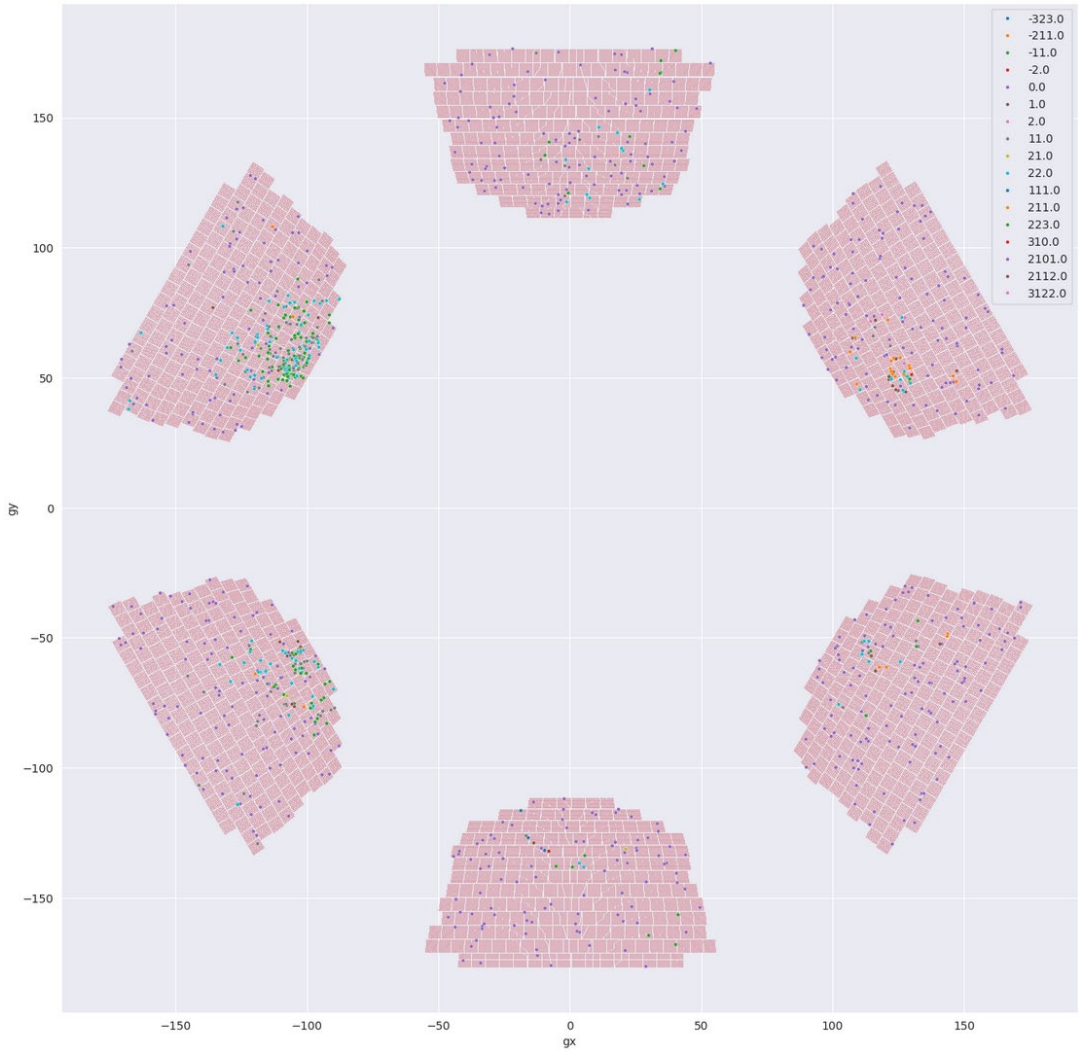

With the new "realistic" noise model, the **global density of dark count noise hits** within the selected events has increased with respect to the previous dataset used for training and testing.

⇒ we decided to maintain the same NN model architecture (in order to cope with the hardware constraints) and to re-train the whole model with the new dataset

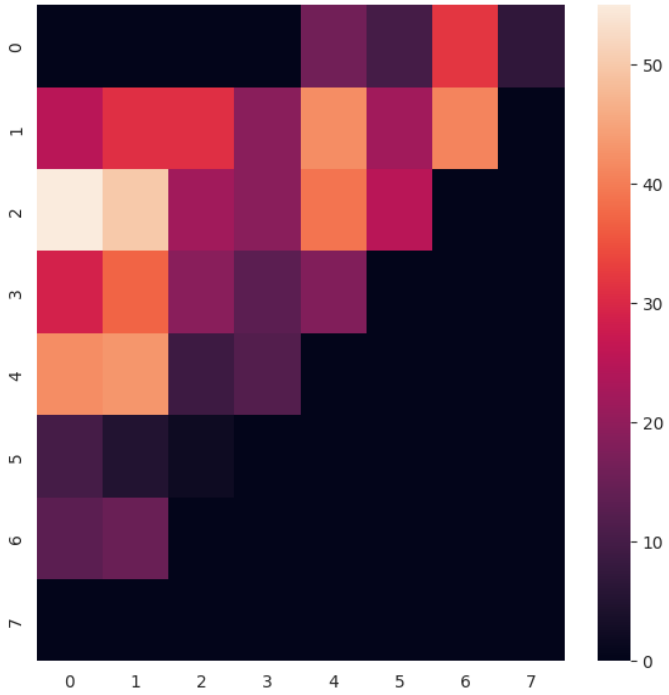


# Updates to the Data Reduction Pipeline

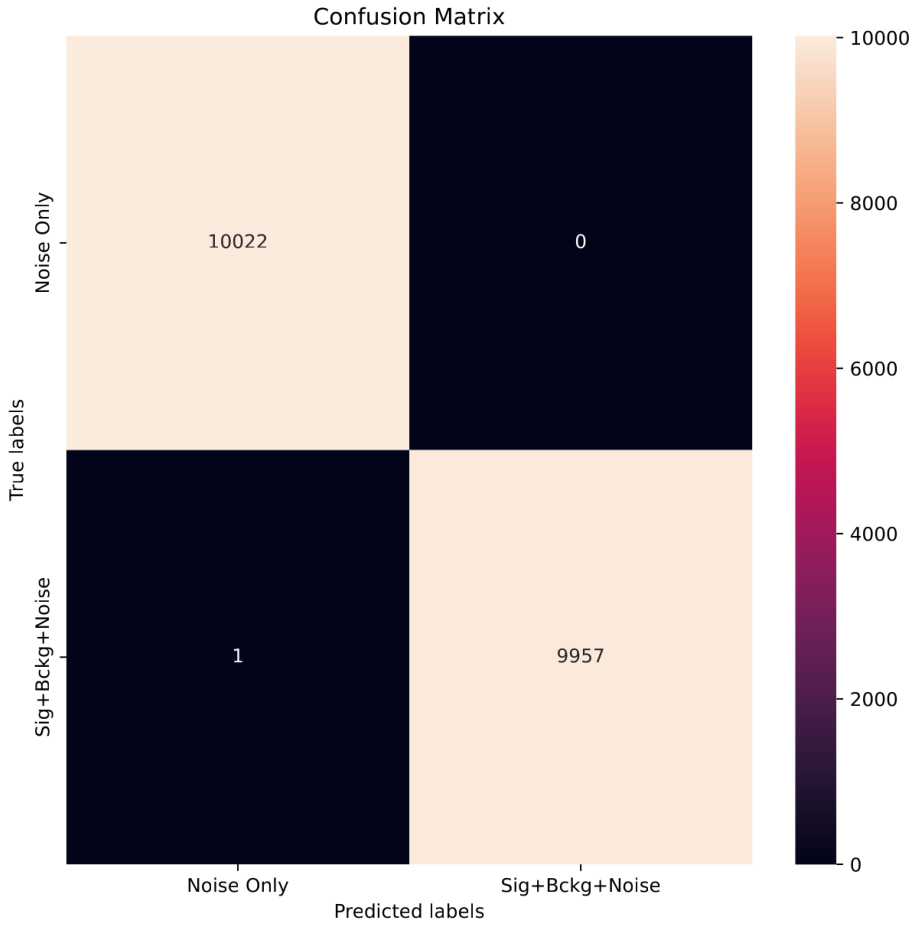



⇒ taking as example a **background event**, we can see how the **pixel intensity in input to the MLP_DAM NN** has increased wrt to the previous cases

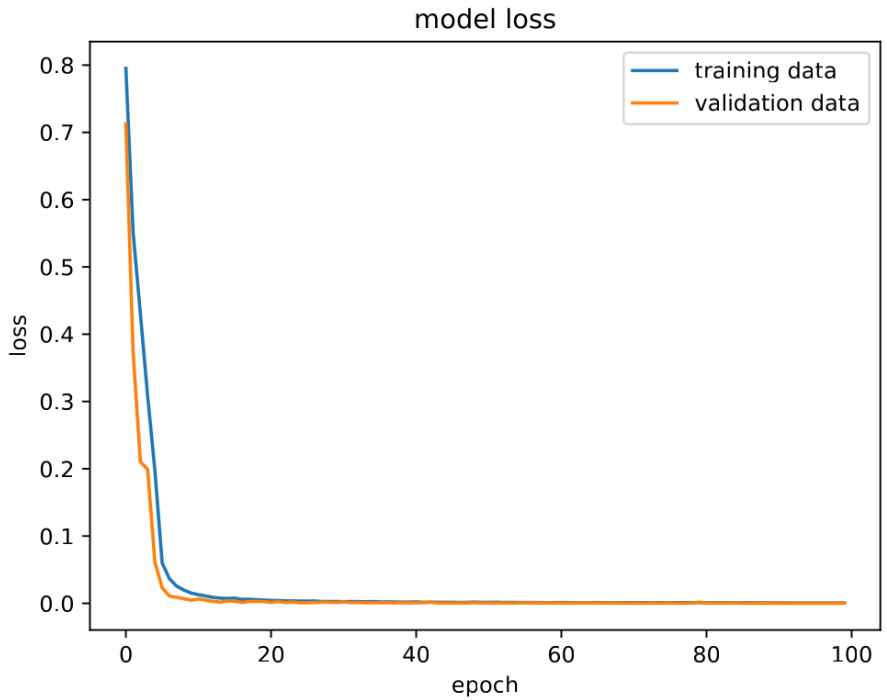


⇒ we explore several new preprocessing steps (rescaling, normalization,...) to take care about this new feature in the data

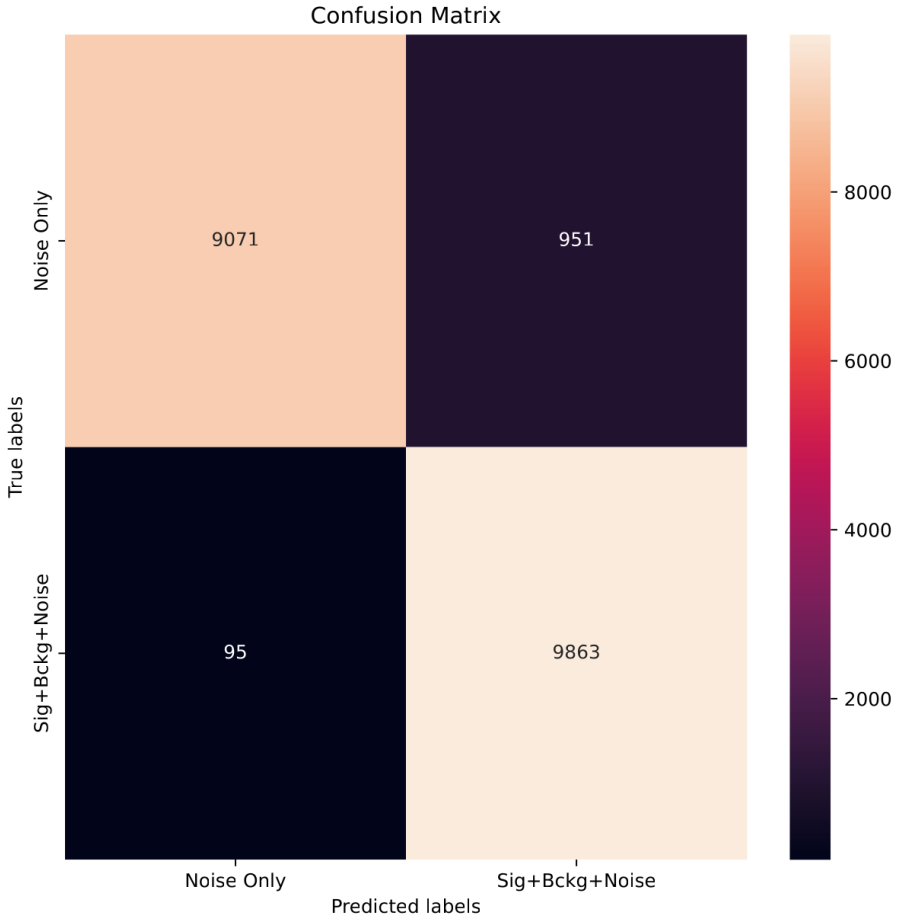
# Results with the new noise model & pipeline



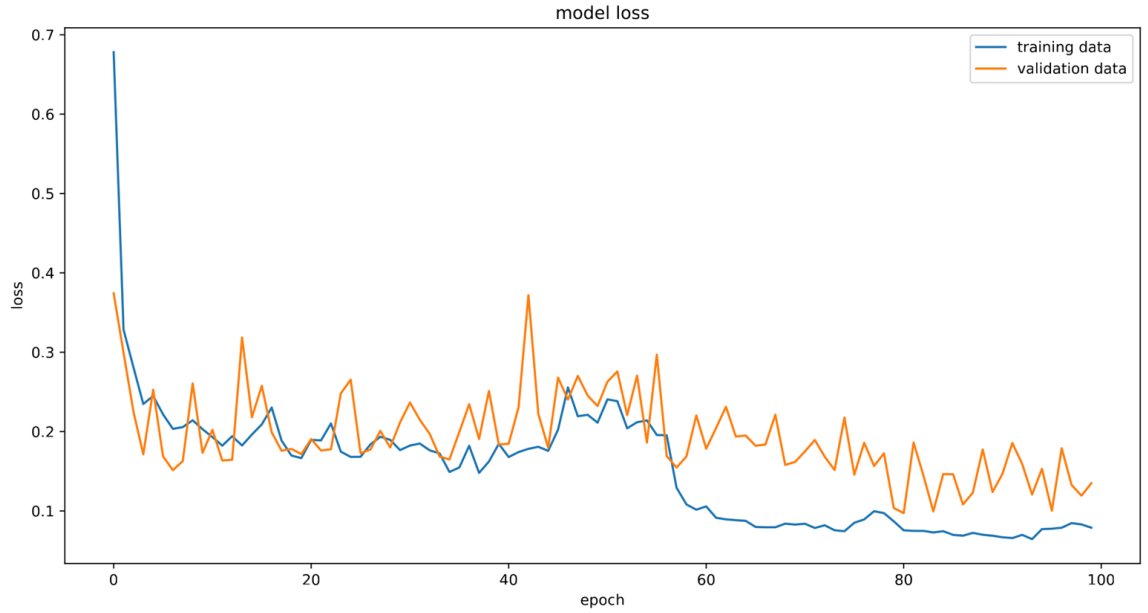

- Accuracy = (TP+TN)/(TP+TN+FP+FN) = 0.9999
- Purity = TP/(TP+FP) = 0.9999
- Efficiency = TP/(TP+FN) = 1.0000



# Results with the new noise model & pipeline



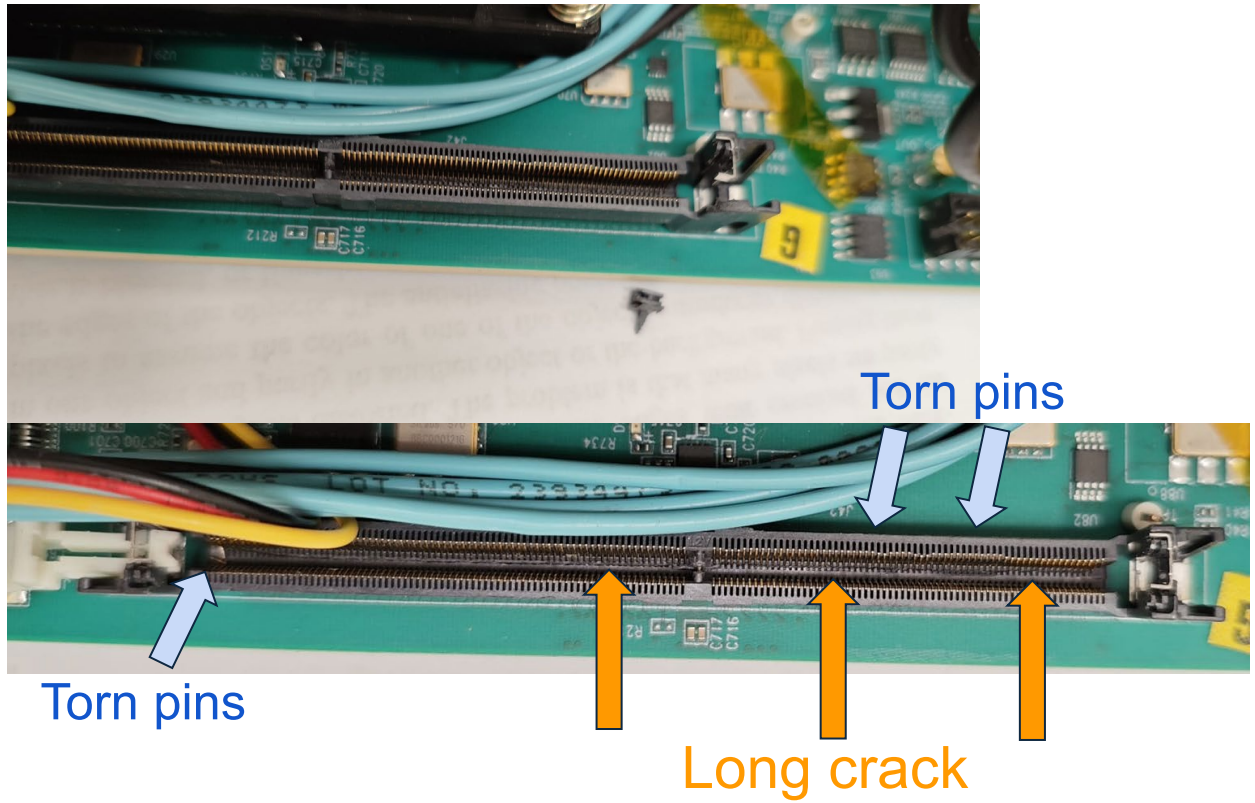

- Accuracy = (TP+TN)/(TP+TN+FP+FN) = 0.9051
- Purity = TP/(TP+FP) = 0.9896
- Efficiency = TP/(TP+FN) = 0.9051




**Model Quantization:**

- Inputs, Activations: fixed point<16,6>
- Weights, Biases: fixed point<8,1>

# Procurement of one Felix-182 Card




Felix-182 board arrived from JLab in Rome end of December ‘24

Unfortunately it showed damages to the DRAM slot:

- **torn contact pins**
- **a crack along the inner side of the slot toward the FPGA**

Maybe they occurred due to the pressure of a DRAM module left in the slot during the delivery (bad packaging).

As a consequence, DRAM is not detected by the system.

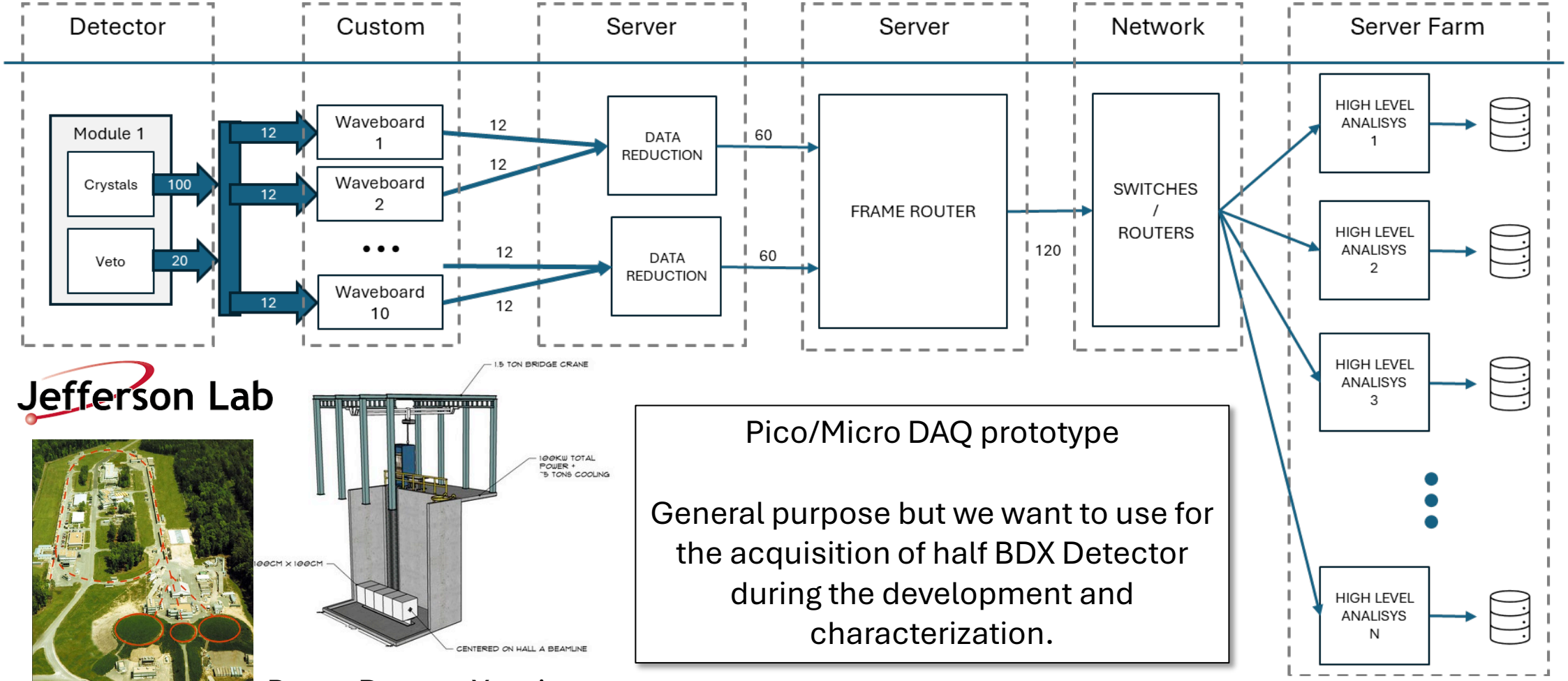
We tried to load some standard fw to the board but it was not possible because the DRAM check failed.

We consulted with:

- the INFN Electronics Laboratory in Rome
- CERN EP-ESE Electronic Systems for Experiments (thanks to Markus Joos)

**Both concur that is extremely difficult to repair the DRAM slot and not worth the effort considering the costs and the likely not optimal result, we are sending it back.**

# Activities in Genova




Jefferson Lab






Beam Dump eXperiment

Pico/Micro DAQ prototype

General purpose but we want to use for the acquisition of half BDX Detector during the development and characterization.

# Organizational contribution

SPADI Alliance

Signal processing and data acquisition infrastructure alliance

Jefferson Lab

Brookhaven National Laboratory

Facility for Rare Isotope Beams at Michigan State University

We all share the Same Goal
Develop a general-purpose streaming acquisition system.

Streaming Data Acquisition System Meeting

**MSU - JLab**

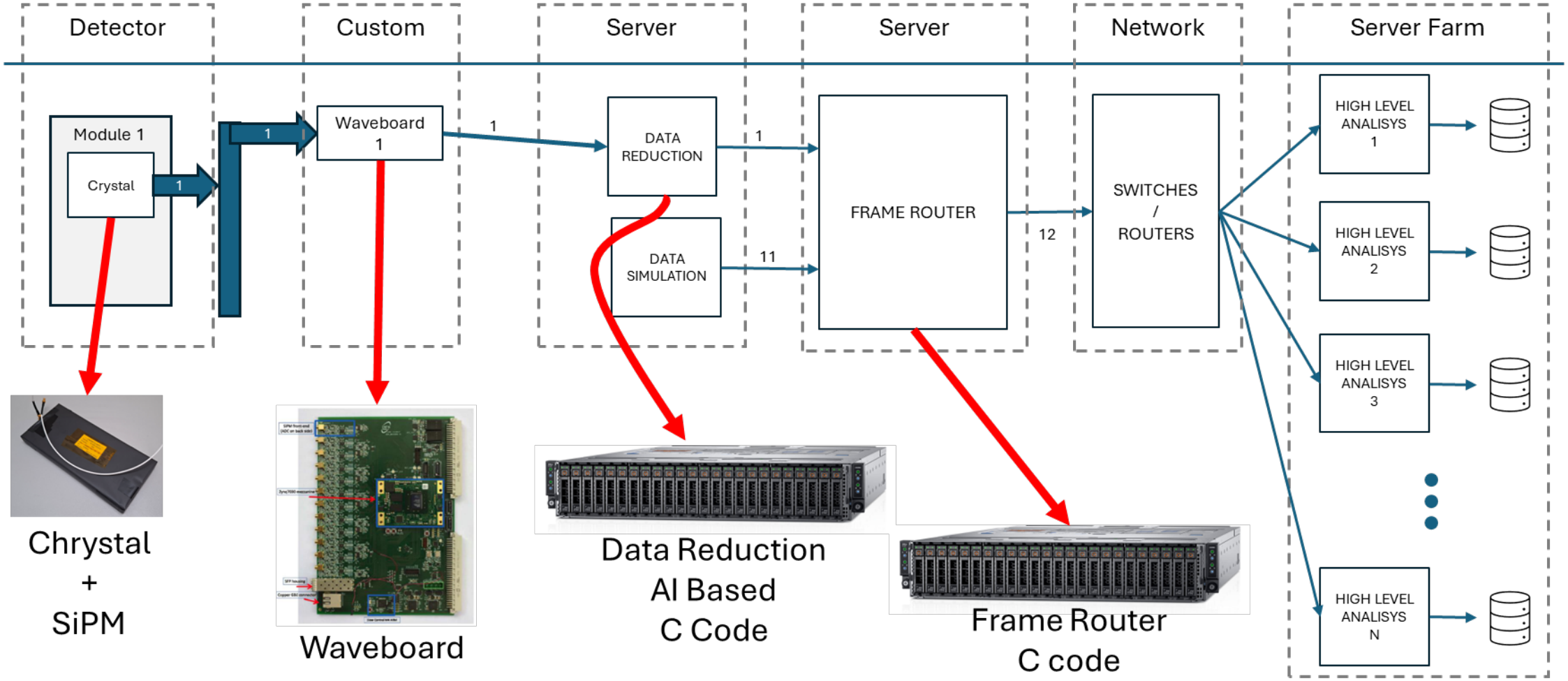
February 12, 2025
Jefferson Lab, Newport News (VA)

Streaming Readout Workshop

**SRO-XII**

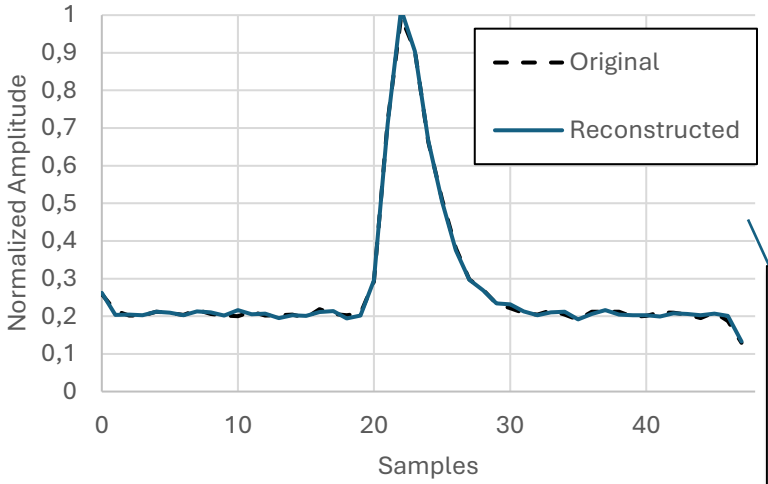
December 2-4, 2024
University of Tokyo, Japan

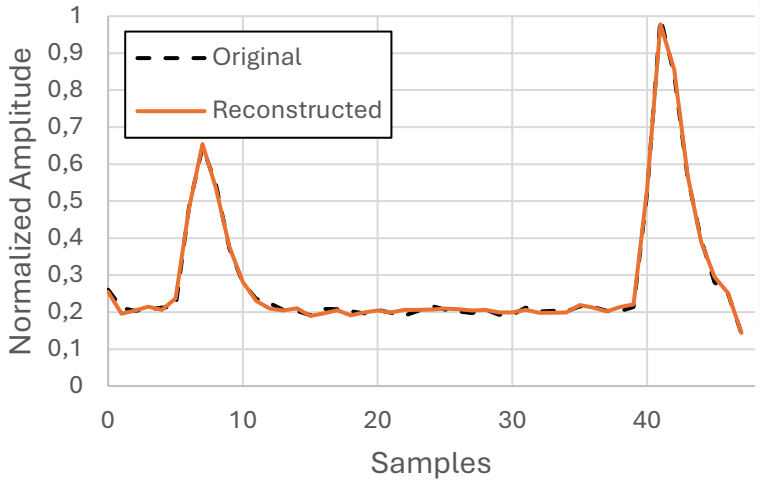
# BDX test Setup: state of work

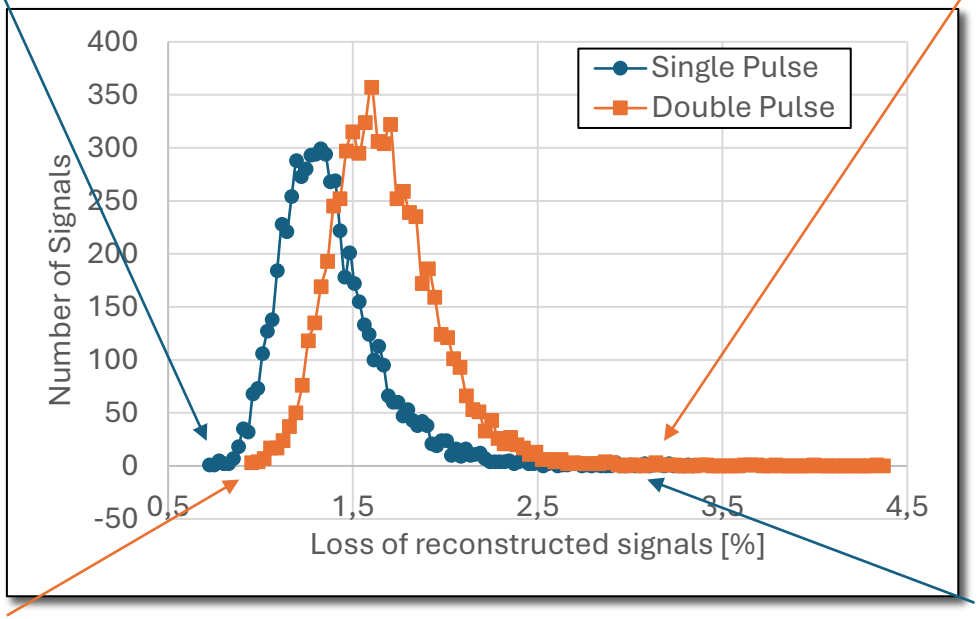


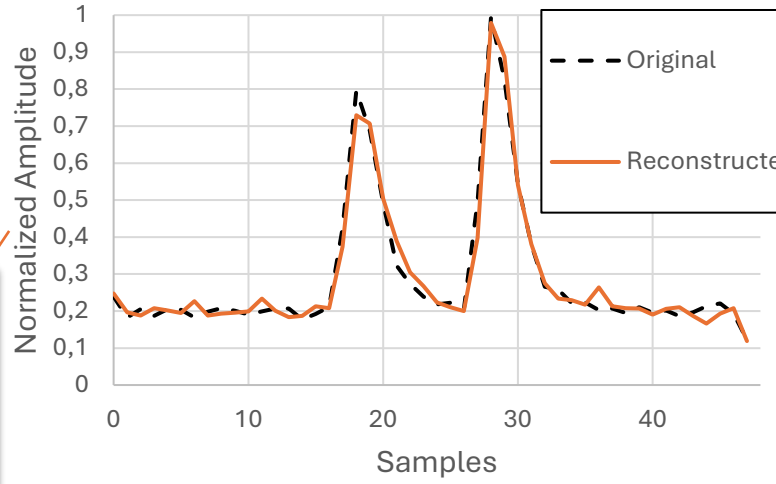
# Data Reduction

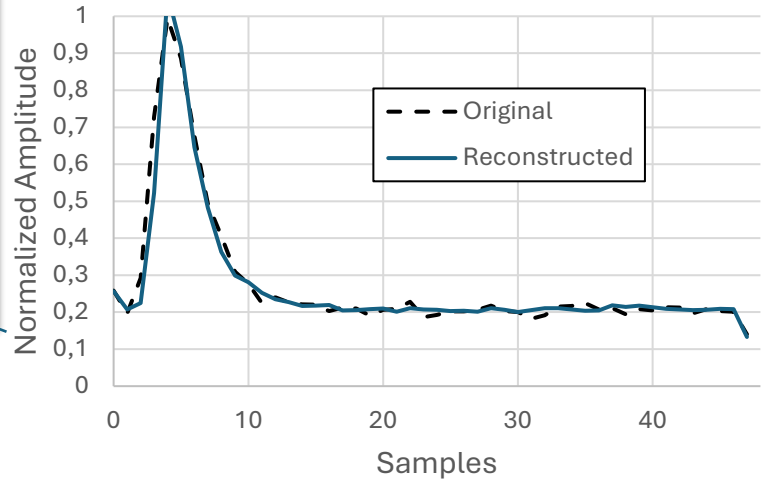
Test results for CLAS12 signals




Lossy compression based on AI algorithm

Compression ratio: 4

Introduced loss: ≈2%

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

# Data reduction: Implementation

NVIDIA Tesla V100 GPU

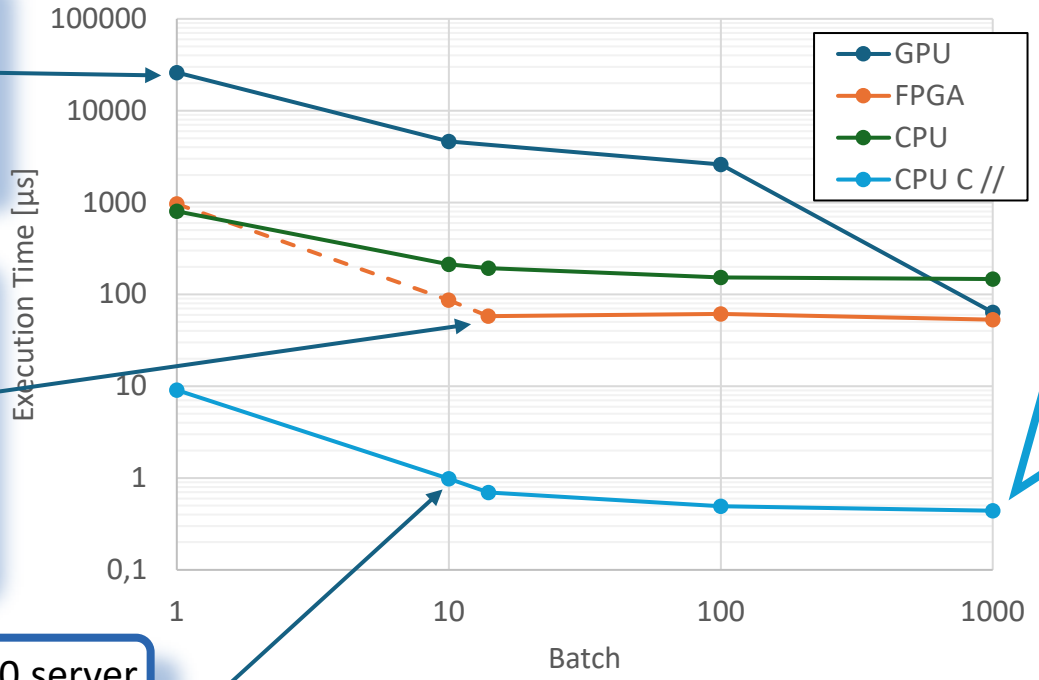
INFN CNAF

Xilinx XRT

XILINX VITIS AI

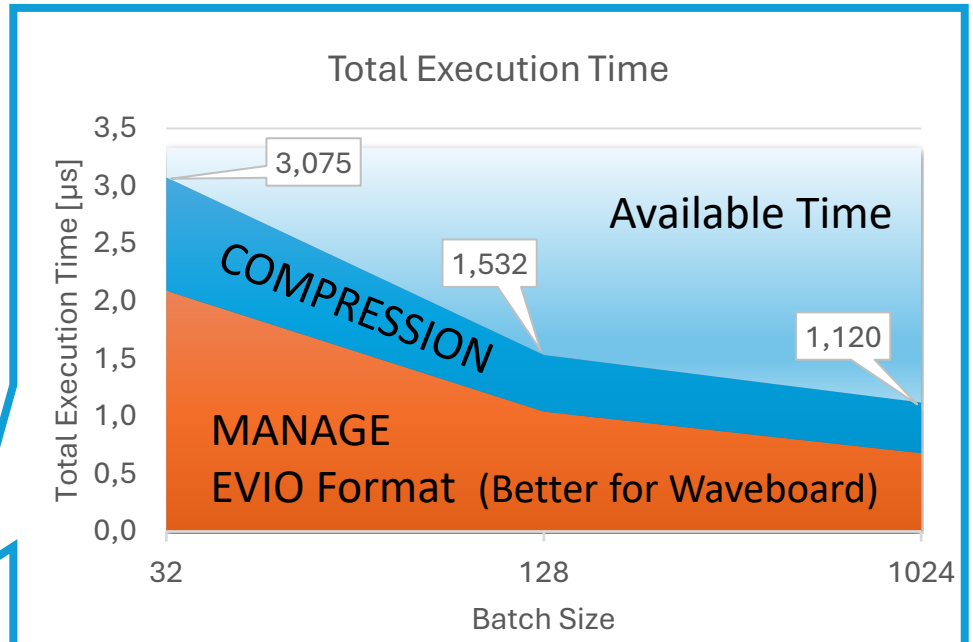
ALVEO V70 FPGA

High performance DELL C6400 server
(4 x AMD EPYC 7413 24-Core)




Test results for CLAS12 signals

AMD presentation @ Workshop on Electronics Torino:
https://agenda.infn.it/event/44098/contributions/253506/



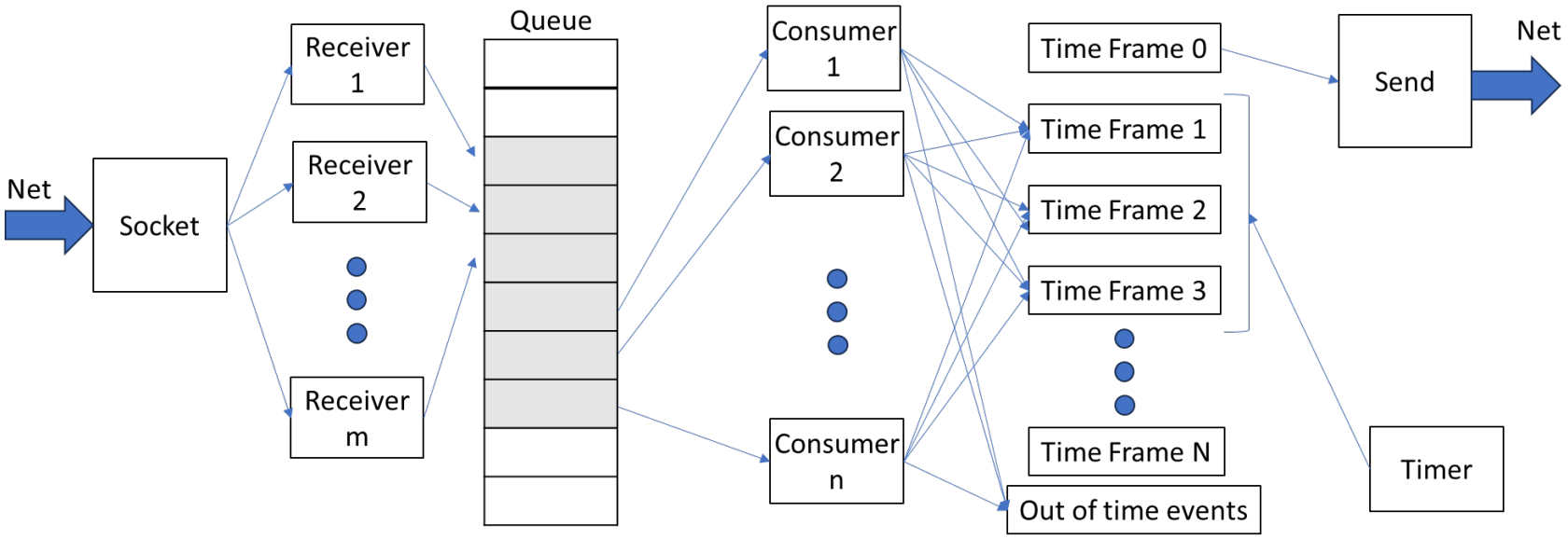

Next Step

Hardware level
HLS4ML
Dedicated connectivity

Alinx VD 100

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

# Frame Router





Details:
- C code
- multithread by means of OpenMP
- CPU Server implementation

First Version are almost done for testing

Required Steps:

Get Full control of Waveboard

Use Multiple Waveboard

Write Custom Run Control

Test and performance evaluation

# Thank you for your attention