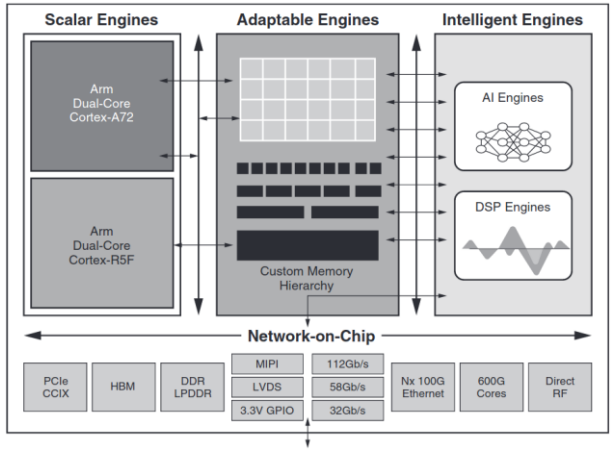
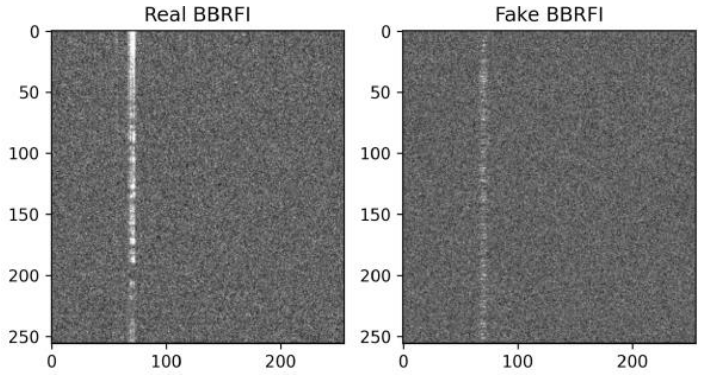
TA5 FPGA Working Group Activities

A. Straessner
for the TA5 FPGA Working Group
PUNCH4NFDI TA5 Meeting
21.09.2023

TA5-XFEL FPGA Workshop

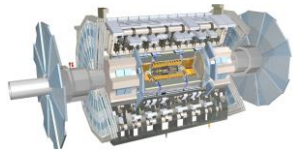
- TA5 - XFEL Machine Learning on FPGA Workshop in June 2023: https://indico.desy.de/event/39436/timetable/?view=standard
  - Invited talks on hls4ml and anomaly detection
  - TA5 contributions:
    - real-time classification of astronomical signals, implementation in hls4ml, deployment on Alveo FPGA cards (MPIfR)
      - artificial signal generation
    - clustering algorithms, physics object detection, evaluation of new FPGA cards including AI engines (Mainz)
      - AI engines not always obvious to optimize to fulfill requirements: latency, ...
    - detector signal reconstruction, ANN implementation, deployment on INTEL FPGA cards (Dresden)
    - deep learning for heavy ion physics (FIAS)
  - Very interesting exchange of experience and solutions with XFEL community
  - Plans to intensify communication (Mattermost, ...) and collaboration:
    - exchange training data
    - exchange ANN models and tool - generalization is a challenge
    - provide concrete tools: FPGA platform for evaluation and training
    - interaction with TA5-XFEL,TA3, Erum-Data, DAPHNE, ...
- Idea of having another workshop in spring 2024



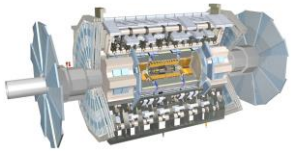


Recent FPGA Working Group Activities

- U. Mainz group:
  - working on alternative architecture for physics object detection in ATLAS forward region - new approach based on CNN layers
  - data sample for CNN training need to be re-processed
  - Bachelor and Master students are part of the team
  - investigate which FPGA types are supported by hls4ml (Versal? Ultrascale?)



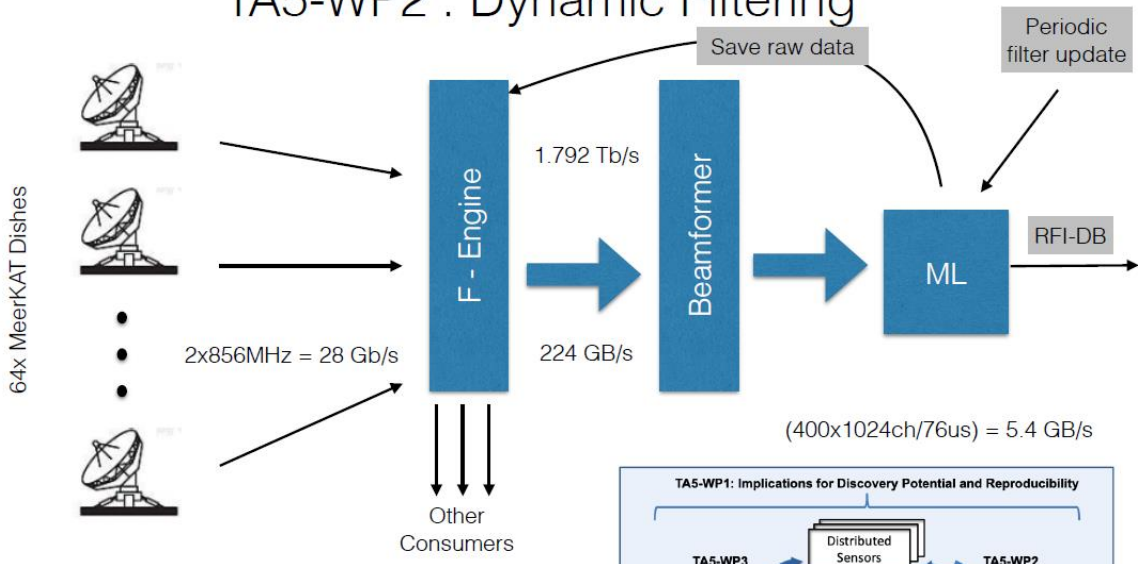
- TU Dresden group:
  - update on application of convolutional neural network CNN application to signal processing
  - improvements in FPGA implementation (VHDL), will also simplify generalization as a firmware tool for other projects



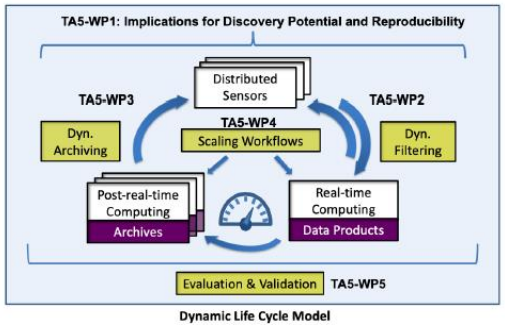
- MPIfR Bonn:
  - tasks now organized in ANN training and FPGA implementation
  - had temporary problems with Alveo FPGA cards which are now solved
  - HLS4ML limits combination number of filters and filter size to 4096; this is to meet timing constraints on FPGA; if networks get larger need to be put on several FPGAs

Recent FPGA Working Group Activities

TA5-WP2 : Dynamic Filtering



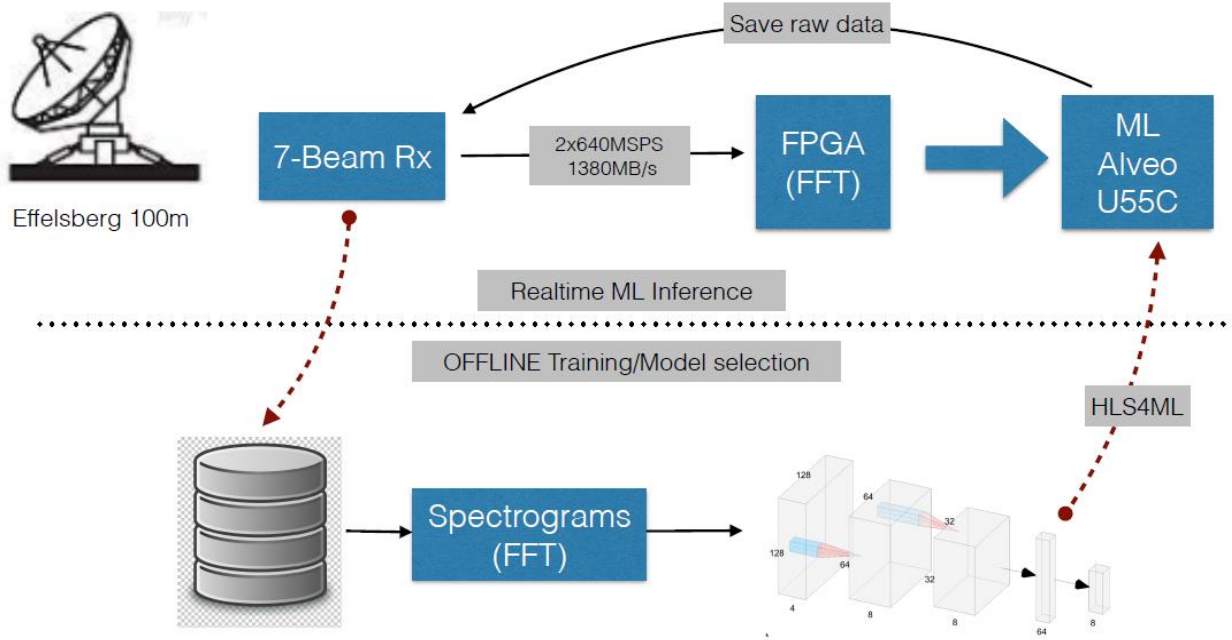
(400x1024ch/76us) = 5.4 GB/s



- up to 960 beam ~ 11GB/s
- SKA is upwards to 2000 beams + many more dishes

TA5-WP2 : Prototype with Effelsberg

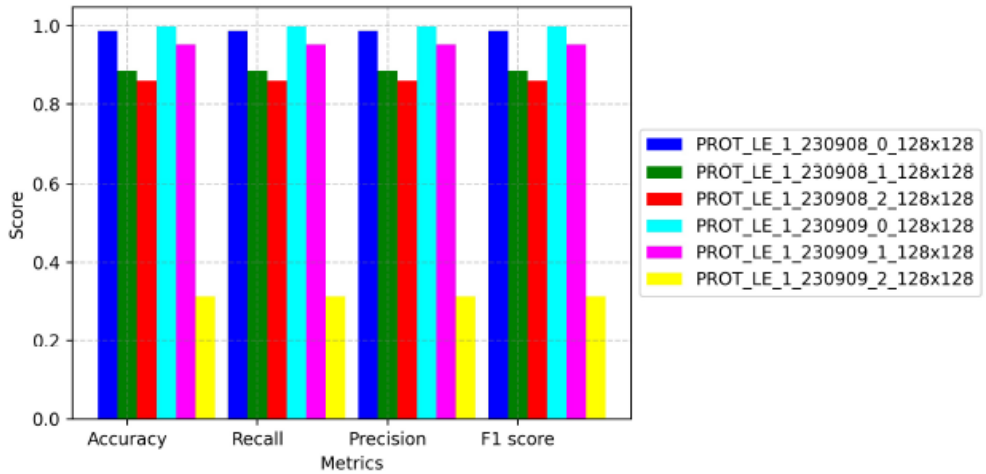
Max-Planck-Institut für Radioastronomie



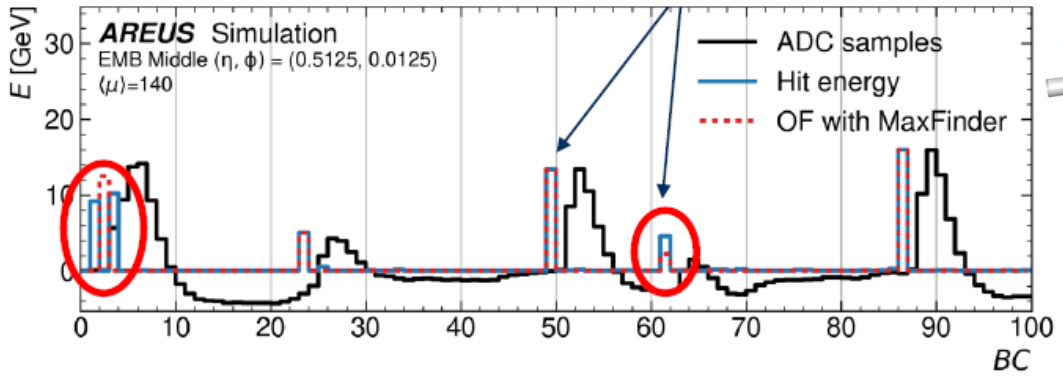
TA5-WP2 : Prototype with Effelsberg

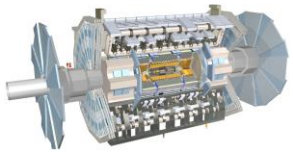
- 20m of ADC output on disk - 1.5625ns sampled, 8-bits
  - contains dispersed radio pulse and RFI
- Data transformed to spectrograms: 256x256, 1.5625MHz, 102.4us = 46150 spectrograms
- Expert-inspected, labelled spectrograms - Training data set with real data
- Training data set augmented with simulated data to cover larger parameter space (DM and S/N)
- TensorFlow based CNN (3 convolutional layers, 4 to 8 5x5 filters, 8 outputs).
- Keras tuner to explore hyperparameters
- Horovod for distributed training - cuts down training by 10x
- Model being targeted to Alveo U55C
  - 1st Version to use high bandwidth memory on U55C to hold spectrograms
  - 2nd Version to have 100GbE interface to stream data from telescope
  - nfilt*filter_w*filer_h*nfilt_last_layer < 4096 : HLS4ML loop unroll limitation.

Performance of 5 different models selected using Keras tuner. All models with 3 conv. layers, and 8 outputs.

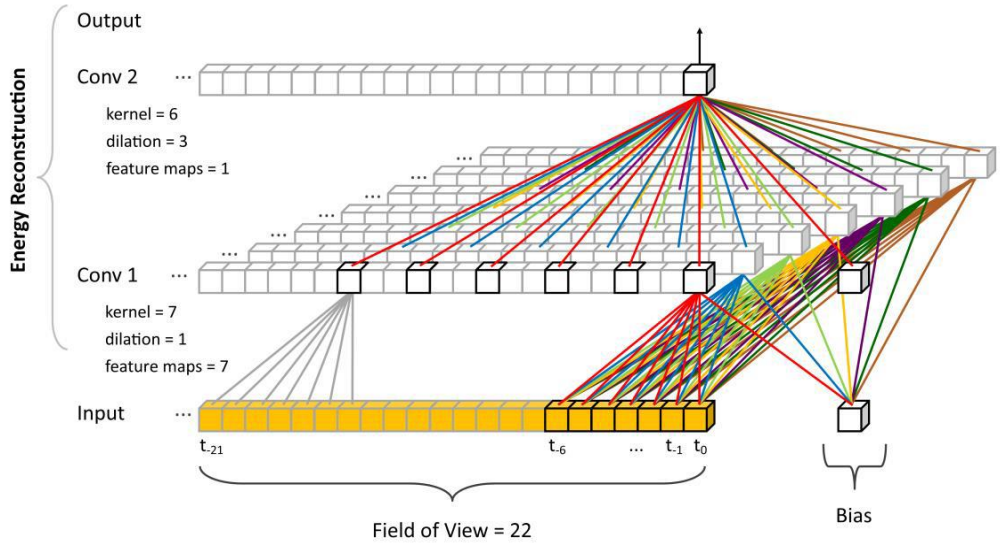


Recent FPGA Working Group Activities

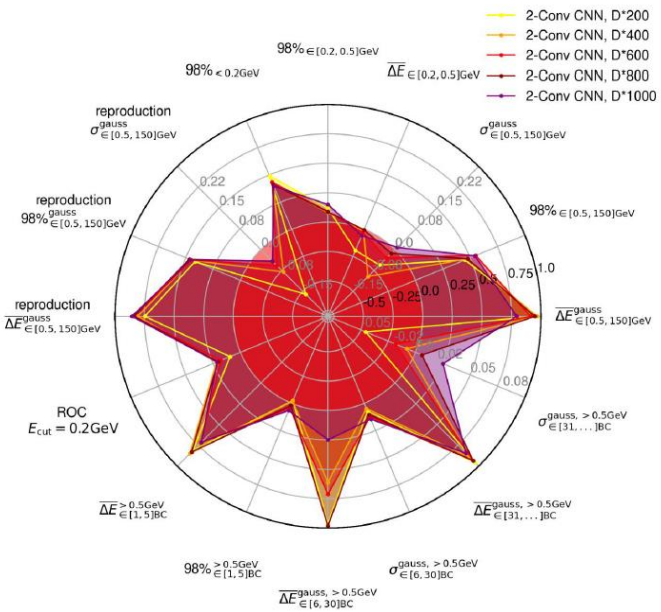




- Development of Neural Networks to improve energy reconstruction



Varying Size of Training Dataset N



Zero gap not resolvable by CNNs

→ Add another dataset to training dataset which holds 0 gaps

Zero gap behavior improved but energy resolution drops