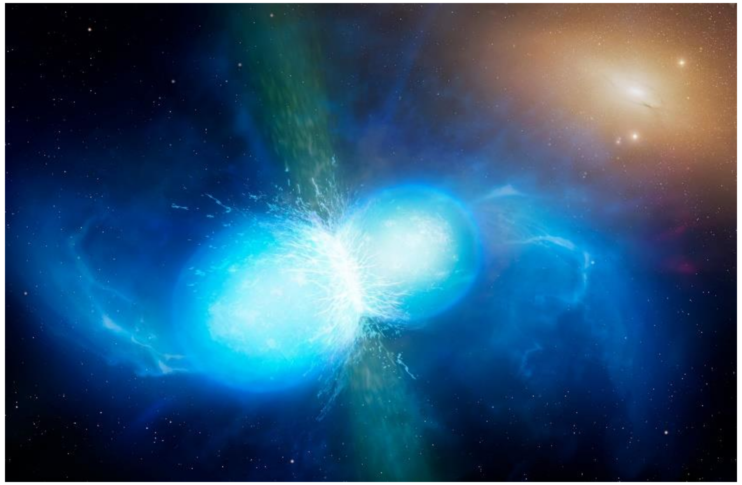
A Neural Network-based trigger for detecting ultra-high-energy neutrinos for RNO-G and IceCube-Gen2

Pawel Marciniewski, Christian Glaser, Alan Coleman, Thorsten Glüsenkamp
Uppsala Univsersity

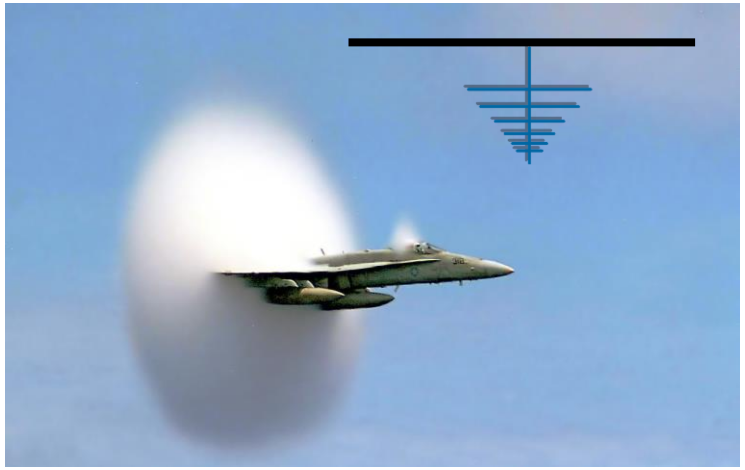
CARL TRYGGERS STIFTELSE
FÖR VETENSKAPLIG FORSKNING
Swedish Research Council
erc
European Research Council
Established by the European Commission
UPPSALA UNIVERSITET

Physics background

- The detection of ultra-high-energy neutrinos would be one of the most important breakthroughs in astro-particle physics in the 21st century.
- It would open a new window to the most violent phenomena in our Universe, e.g., what happens in the vicinity of supermassive black holes, in neutron star mergers, or gamma ray bursts. UHE neutrinos are excellent probes of astro-particle and high-energy physics both within and beyond the Standard model, by e.g. studying their production mechanism.
- Ultra-high energy neutrinos can be detected using the radio emission that they create when interacting in dense media, such as ice. Short radio wave pulses with the power spectrum ranging from tens to hundreds of MHz can be detected with the help of antennas buried in glaciers.



An artistic view of neutron star merge



A Cherenkov radiation cone and an antenna

In-Ice Radio Neutrino Detection Experiment Landscape

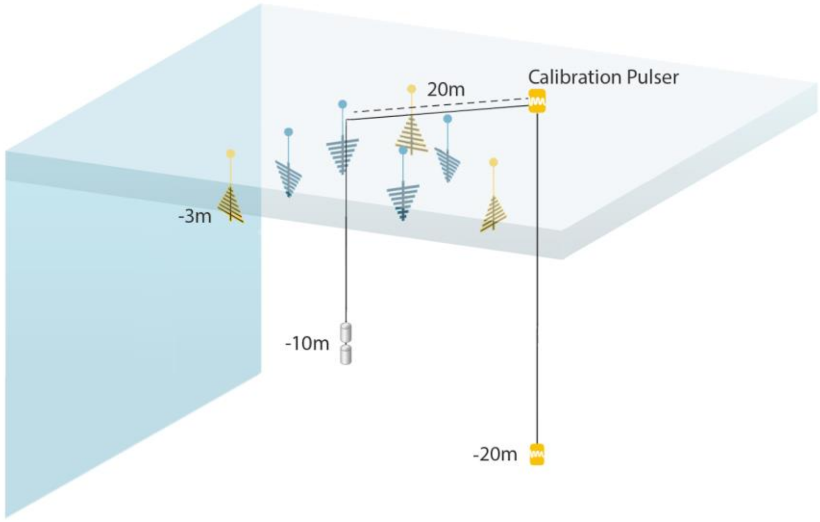
past **now** **future**

ARIANNA test bed

- 12 shallow stations at Moore's Bay + South Pole

ARA

- 5x 200m deep stations at South Pole

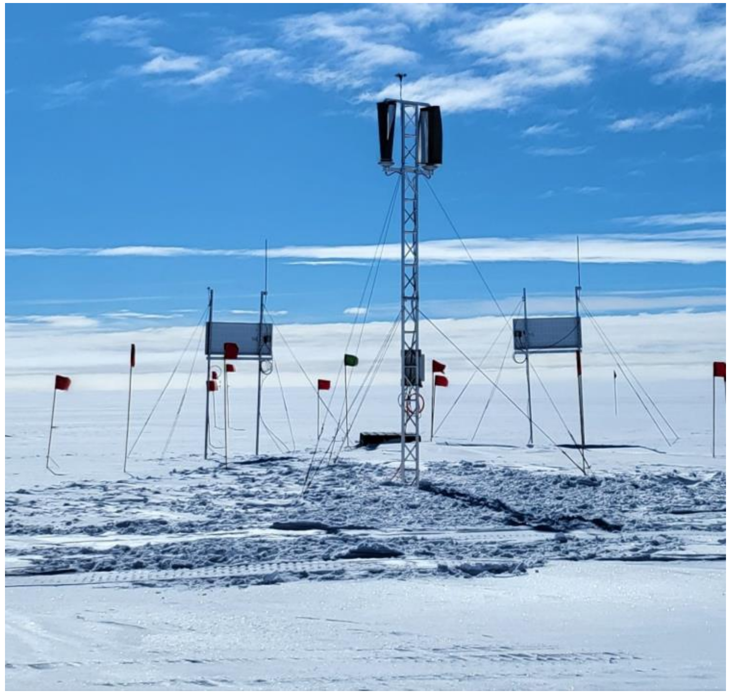


RNO-G

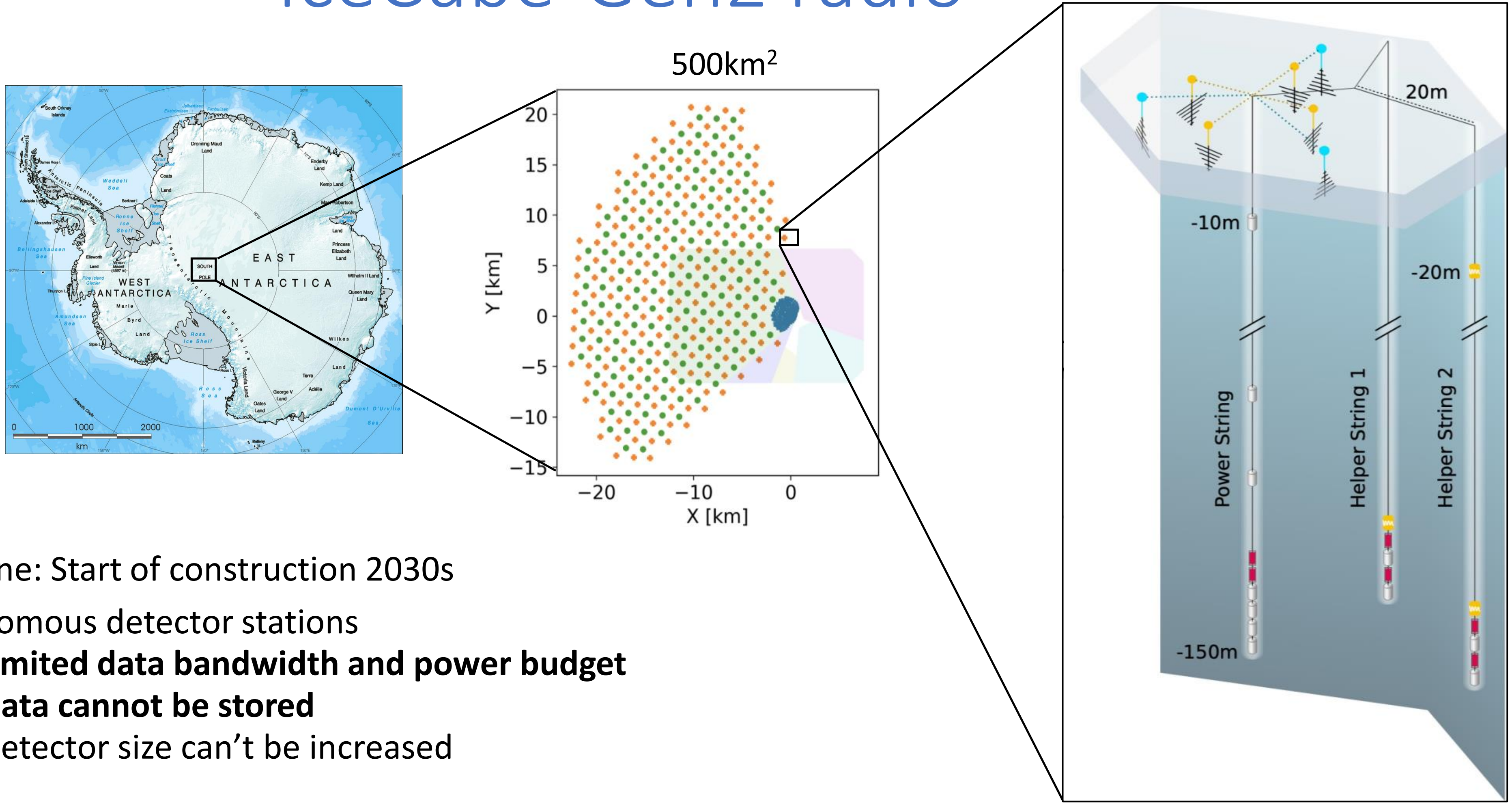
- 35+ detector stations in Greenland
- first deployment summer 2021
- First UHE neutrino discovery likely

IceCube-Gen2

- 300+ detector stations at South Pole



IceCube-Gen2 radio

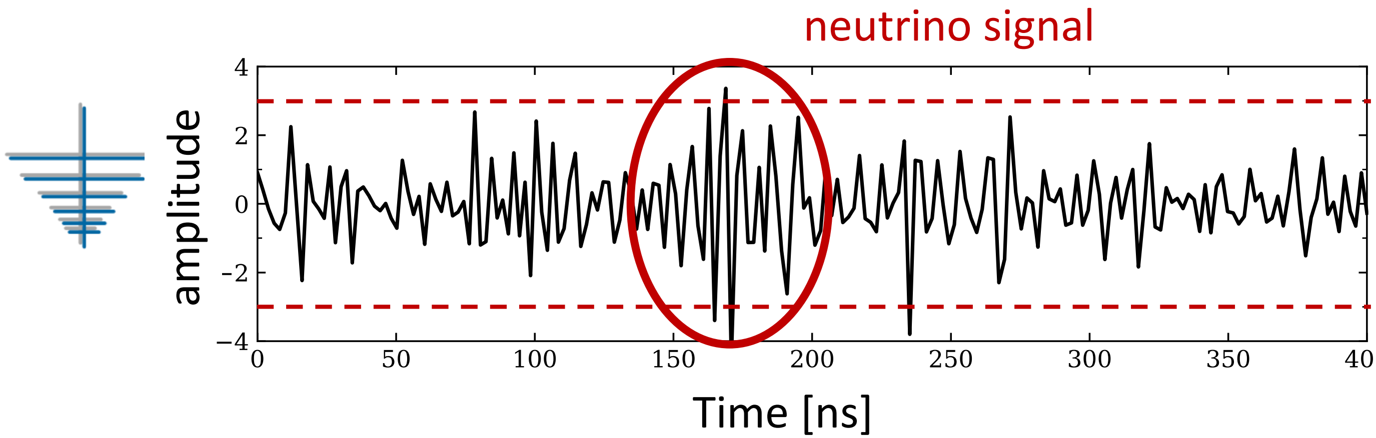


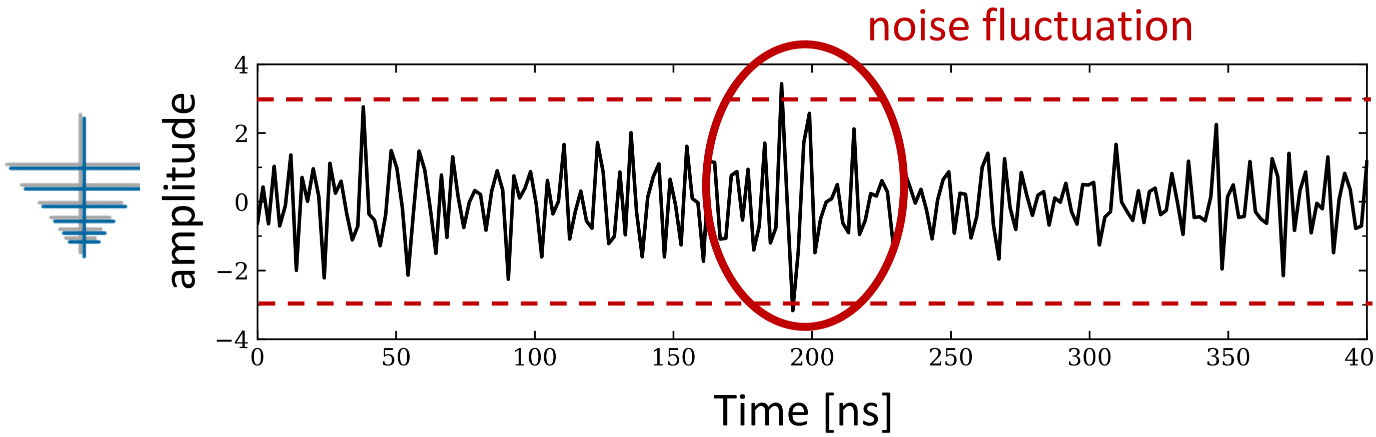
- Timeline: Start of construction 2030s
- Autonomous detector stations
  - **limited data bandwidth and power budget**
  - **data cannot be stored**
  - detector size can’t be increased

→ Only option to accelerate the research field: better detector (this project)

Antenna signal and Triggering techniques

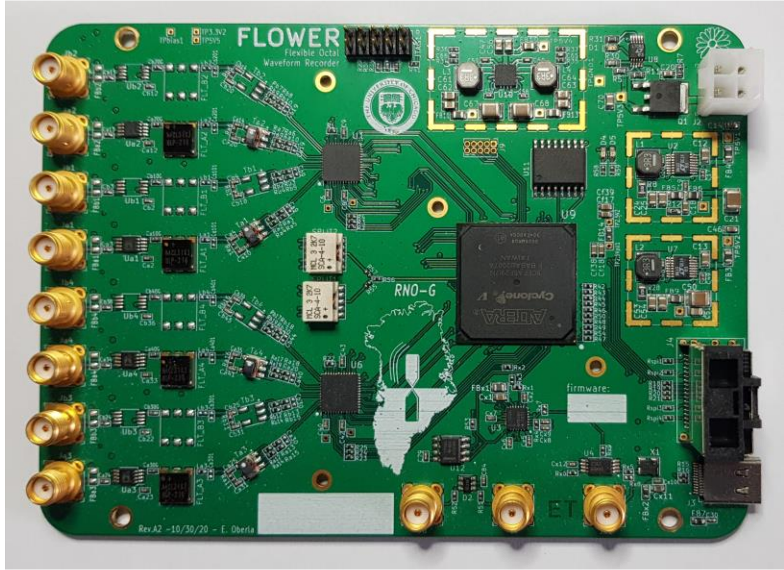
- Threshold-based trigger
  - Noise fluctuations dominate trigger
  - Thresholds need to be high enough to limit trigger rate on thermal noise.
- Coincidence-based trigger
  - Requires simultaneous threshold crossing for several antennas
- Beam forming
  - Performs signal phase shifts on antenna arrays to detect coherent power increase
  - Capable of determining the wave angle
- Deep learning
  - Substantial prevail in terms of efficiency according to simulations.





The goal - step 1

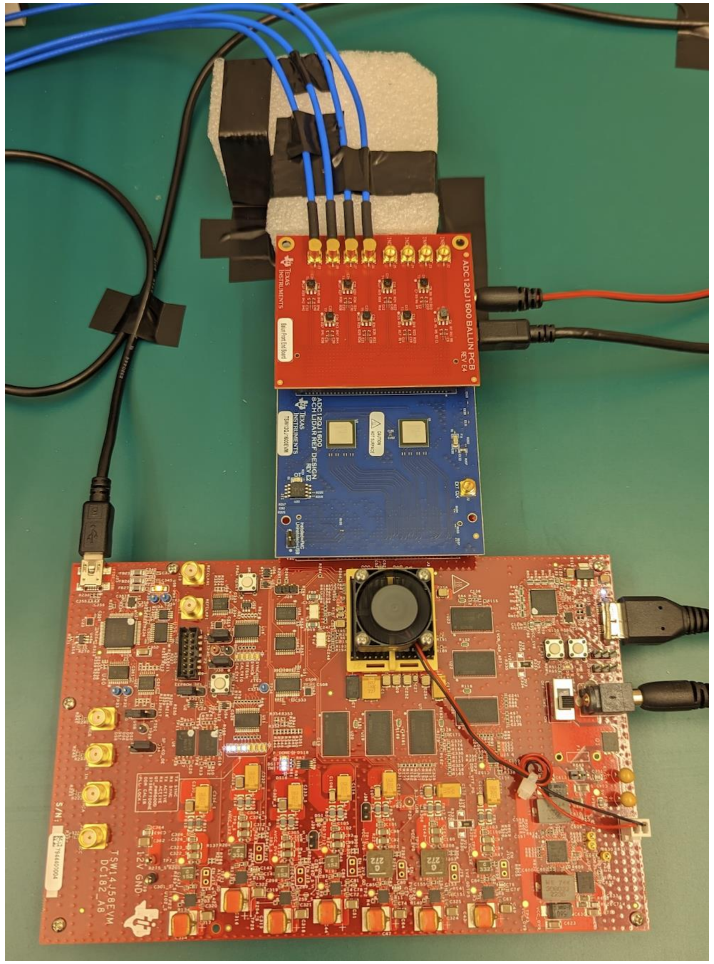
- In this work we present a trigger system utilizing a Convolutional Neural Network to processes the antenna signals digitized by 500 Msps, 8-bit ADCs
- In the first stage we will construct an CNN trigger as a second level, to be pre-triggered by a threshold-based trigger. This design will be implemented on an existing hardware (FLOWER board *) and used for conceptual studies in the RNG-O experiment.



FLOWER - Low Threshold Trigger Board

The goal - step 2

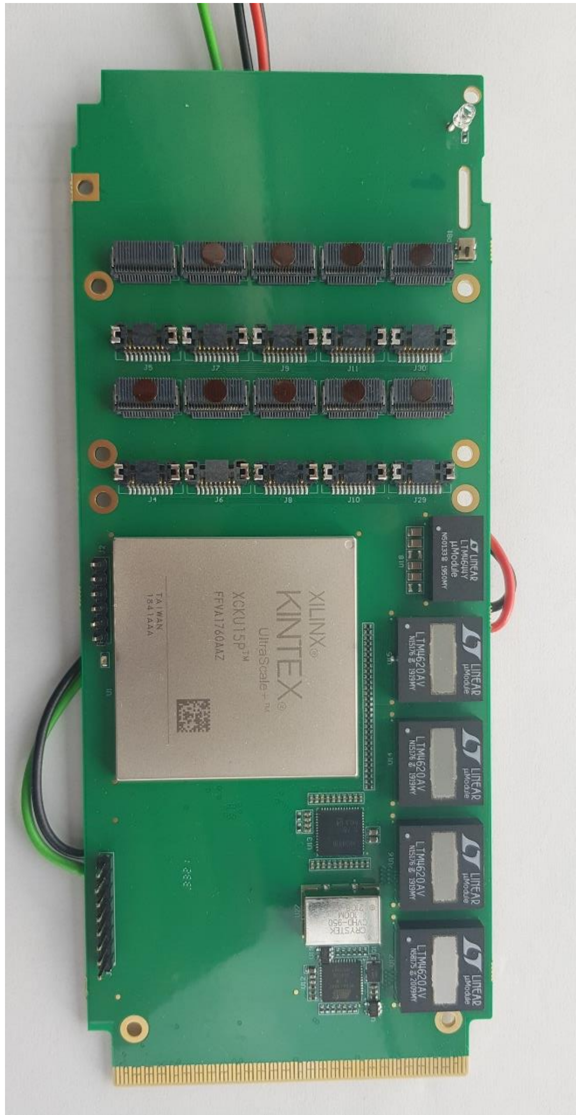
- In the second stage we aim at implementing the CNN trigger as a main trigger on a future DAQ system, comprising 4-channel, 12-bit, 1 Gsps ADC and a Kintex UltraScale+.
- The customized hardware will be built basing on experience from an evaluation platform.
- The system is intended for the IceCube-Gen2 radio experiment
- Well-trained system will detect sub-threshold events



1 Gsps evaluation platform

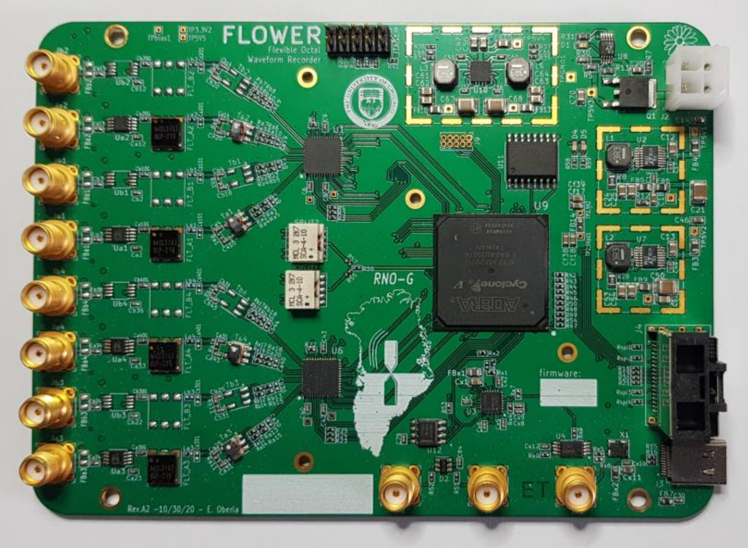
Firmware evaluation environment

- Evaluation of firmware to be implemented on the FLOWER board we use a popular evaluation board from Terasic (DE1-SoC), comprising an Intel/Altera Cyclone V FPGA.
- The evaluation of firmware for the future DAQ we use a prototype of a PANDA-DC (an AMC-based DAQ board designed for the PANDA experiment comprising an AMD/Xilinx Kintex Ultrascale+ and 60 optical links



PANDA-DC prototype

RNO-G Phased Array (Beamforming) Trigger

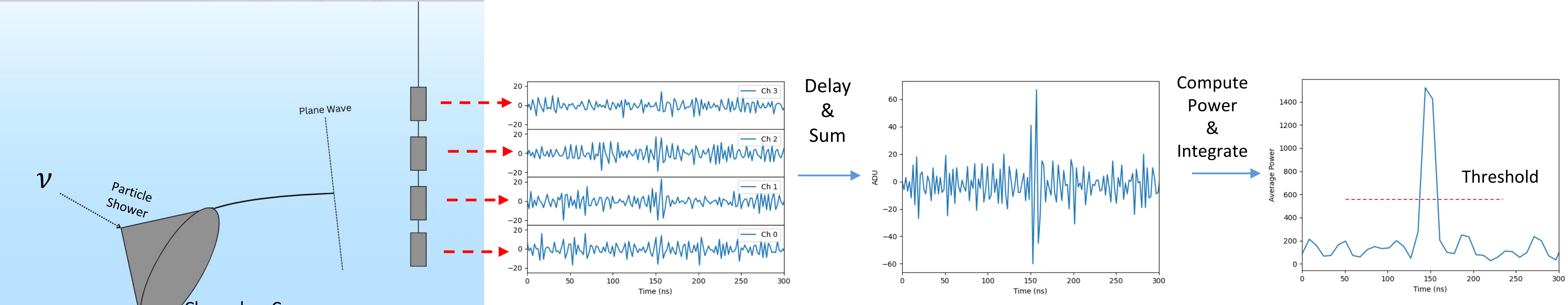


FLOWER - Low Threshold Trigger Board

- 4 Channels using 2x HMCAD 1511 Streaming ADCs to a Cyclone V FPGA
- 8 Beams spaced equally from elevation angles of -60° to 60°

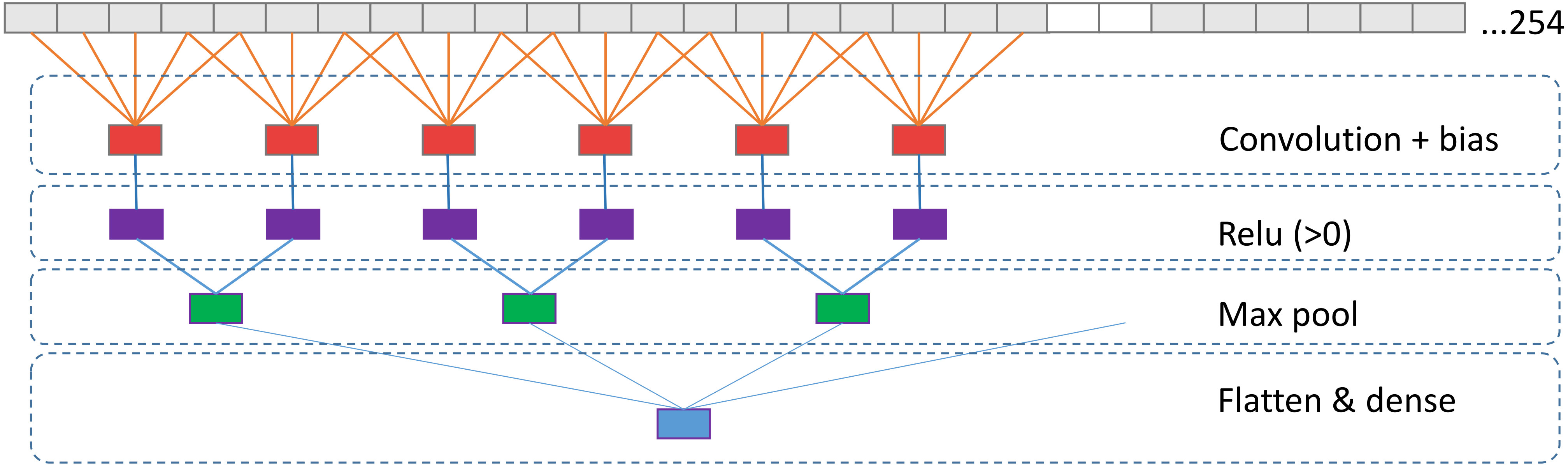
Trigger Calculation

- 4x Linear Interpolated Samples (low latency & simple) for sub-sample beams
- For each beam, delay and sum waveforms based on arrival times
- Calculate power of the samples (squaring the samples)
- Integrates (averages) the instantaneous power over 16 ns
- Average Power is compared to a threshold for a trigger



Thanks to Ryan Krebs

The Neural Network concept



...254 samples

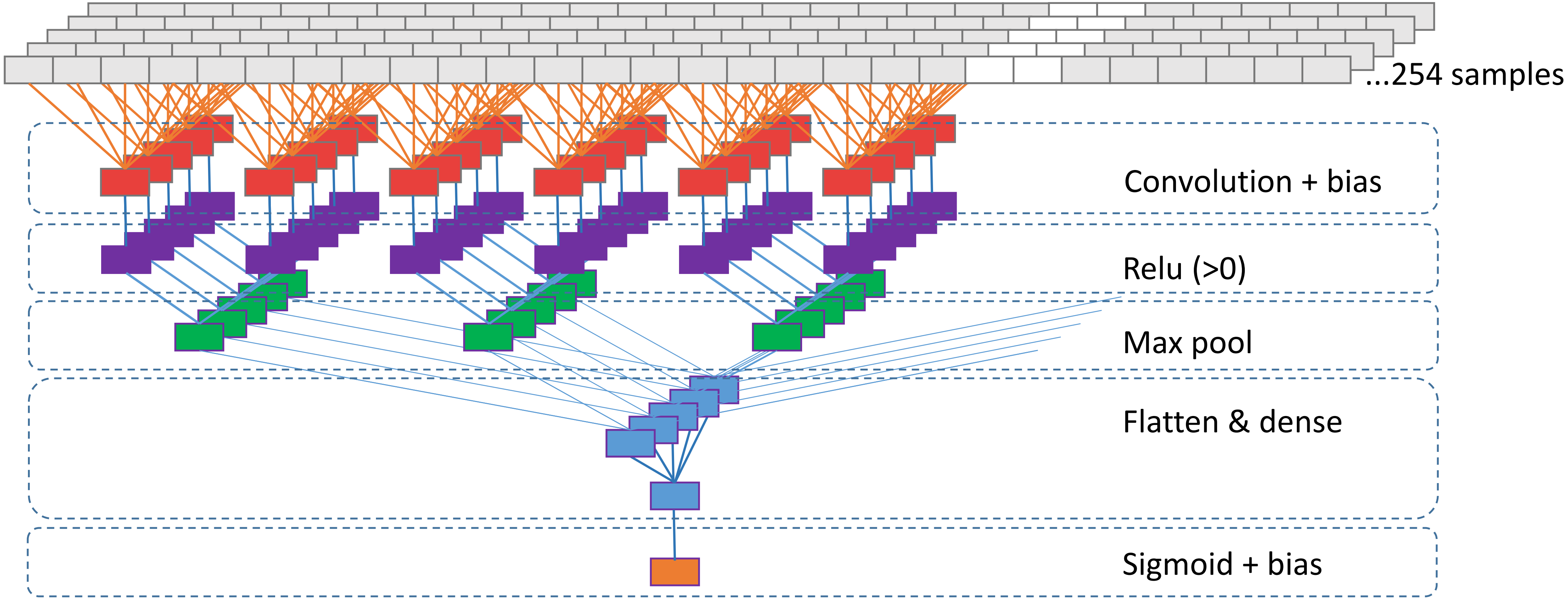
Convolution + bias

Relu (>0)

Max pool

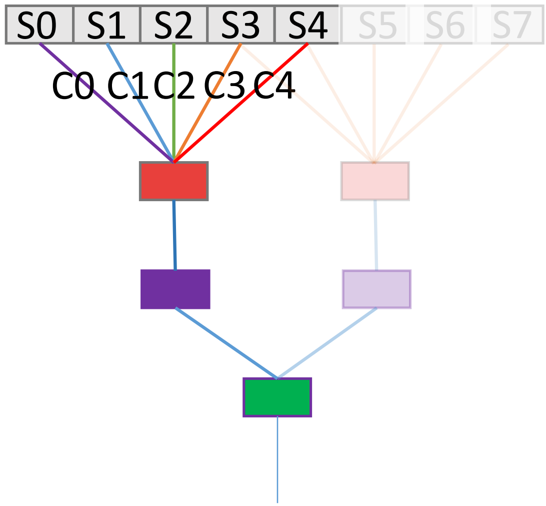
Flatten & dense

The Neural Network concept



The Neural Network concept

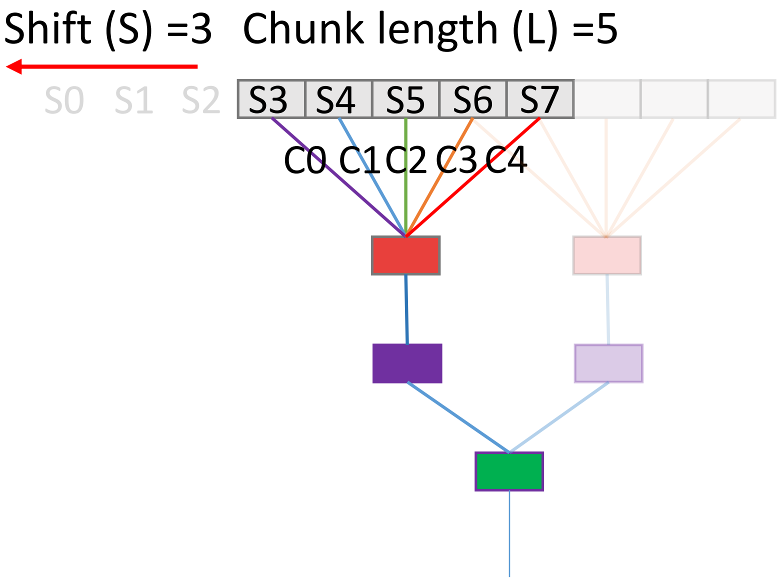
Chunk length (L) =5



Observations

1. Data are streamed by ADCs. Data processing may start with the first input word and proceed without waiting for aquiring of the whole record.
2. The five convolution coefficients set is the same for all multipliers. This simplifies modularization.
3. Using muliply-accumulate is the optimal way of data processing since one uses a dedicated hardware. Arguably It would hardly be faster to use 5 parallel multipliers and then subsequently add partial results.
4. The data chunk needs to be frozen for the processing of it’s contents and then a subsequent chunk is moved in

The Neural Network concept

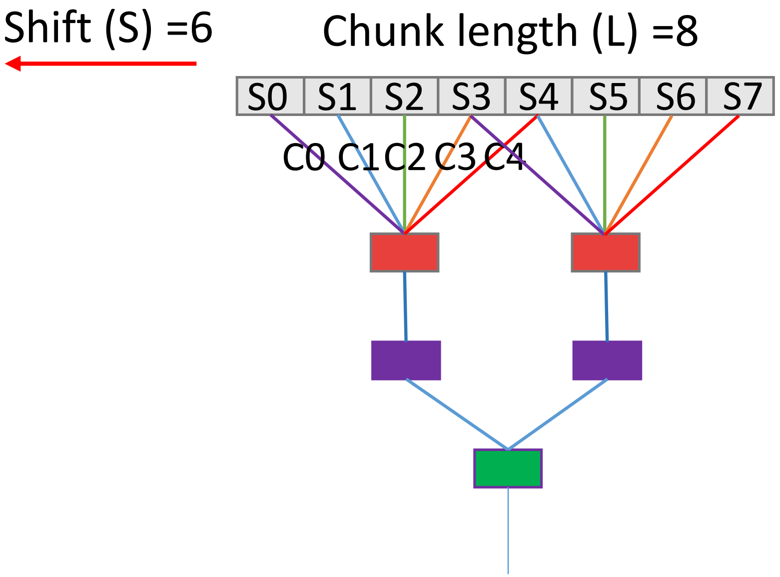


Shift (S) = Length (L) - 2

Observations

1. Data are streamed by ADCs. Data processing may start with the first input word and proceed without waiting for aquiring of the whole record.
2. The five convolution coefficients set is the same for all multipliers. This simplifies modularization.
3. Using muliply-accumulate is the optimal way of data processing since one uses a dedicated hardware. Arguably It would hardly be faster to use 5 parallel multipliers and then subsequently add partial results.
4. The data chunk needs to be frozen for the processing of it’s contents and then a subsequent chunk is moved in

The Neural Network concept



Observations

1. The processing can be faster if it's performed by more multipliers simultaneously
2. For continous data processing:
   the processing time of the record $\boldsymbol{T_M}$ must be shorter than the record streaming time $\boldsymbol{T_S}$

$$\boldsymbol{T_M} = \frac{N}{M}\tau_m$$

Number of multiplications per chunk

Multiplication time

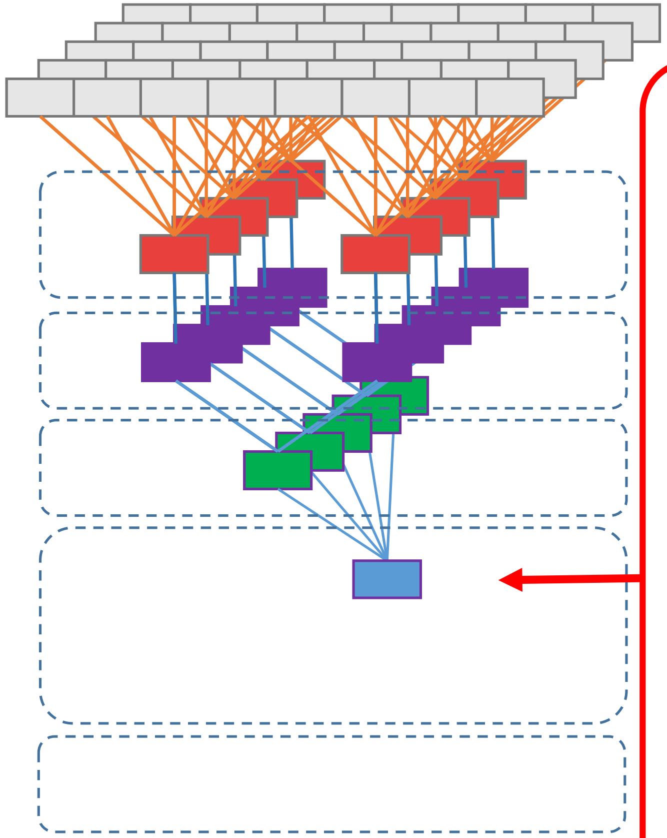
Number of multipliers per chunk

$$\boldsymbol{T_S} = S \cdot \tau_s$$

Shift

Data sampling period

The Neural Network concept

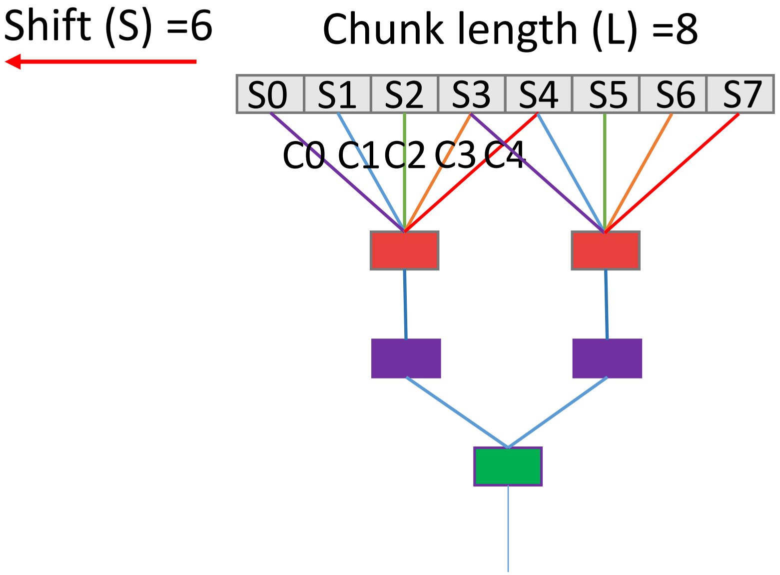


Observations

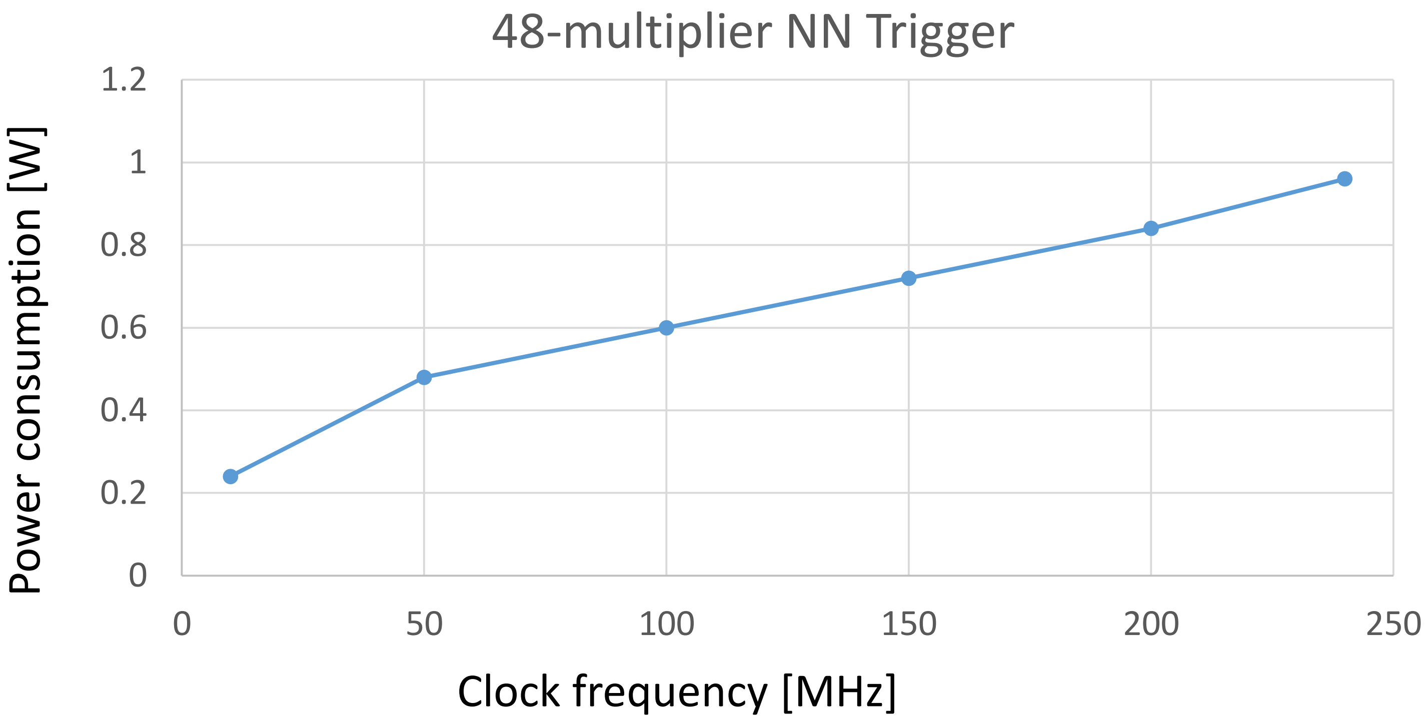
1. The densing stage can also be incorporated the core and perform on-line multiplication/accumulation of 5 patterns
2. The work load is then well balanced between stages.
3. The densing stage uses 18x9 multipliers and therefore performs slower than 9x9 multipliers in the convolution stage. To solve the bottleneck we are using two multipliers working in interleaved mode.

4. The sigmoid function proves not to be necessary and it’s reduced to a bias

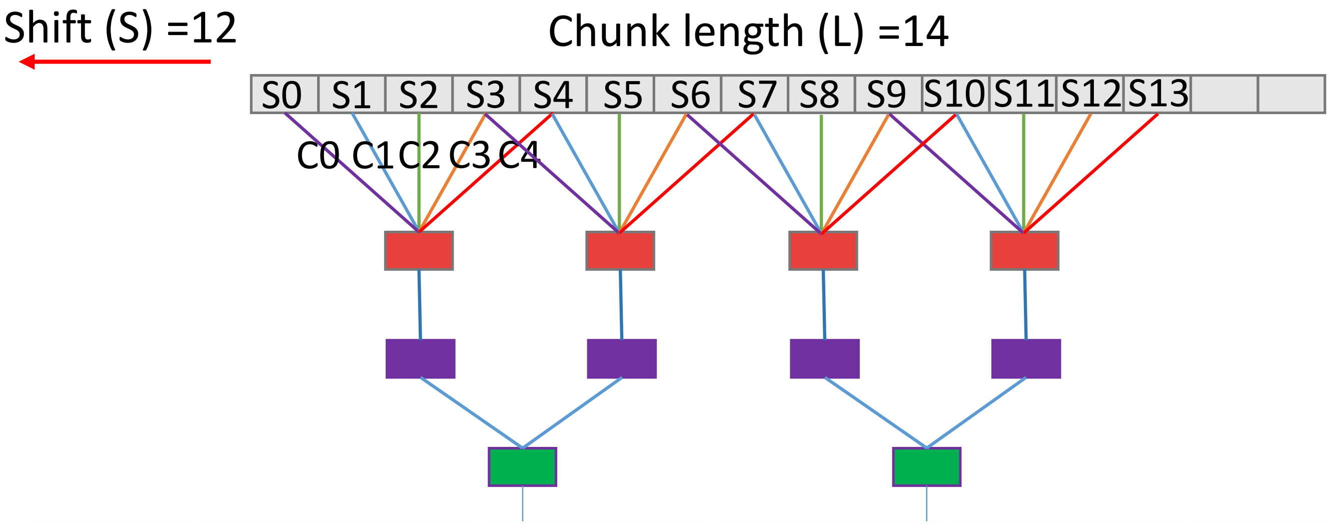
CNN power consumption



Performance check of the multipliers in the non-clocked chain: **Cyclone V**



The CNN implementation



Chunk length L	Number of multiplication N	Number of multipliers M	Shift S	Processing time T_M	Streaming time T_S
5	5	1	3	$5\ \tau_m$	$3\ \tau_s$
8	10	2	6	$5\ \tau_m$	$6\ \tau_s$
14	20	4	12	$5\ \tau_m$	$12\ \tau_s$
20	30	6	18	$5\ \tau_m$	$18\ \tau_s$
26	40	8	24	$5\ \tau_m$	$24\ \tau_s$

Results:

Implementation of the algorithm with chunk length L = 8 allowed for obtaining the following maximum clock frequencies:

Altera Cyclone V – 230 MHz ($\boldsymbol{\tau_m}$ **= 4.34 ns)**
Kintex Ultrascale+ – 230 MHz ($\boldsymbol{\tau_m}$ **= 4.34 ns)**

Conclusions:

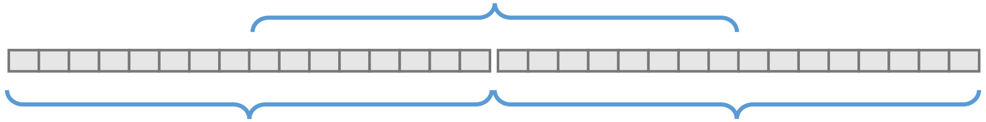
For possible continous on-line processing (triggering) one can use the following options:

FLOWER
Altera Cyclone V & 500 MSPS ADC

Future
Kintex Ultrascale+ & 1 GSPS ADC

The CNN implementation

To solve border issue between records and maintain full detection effficiency it requires simultaneous processing of overlapping data records.



This leads to doubling of the necessary resources

Possible configurations

Technology [ADC clock/FPGA]	Clock frequency [MHz]	Number of multipliers /available	Processing time [us]	Estimated power [W]
500 Msps/Cyclone V	200	48	1.3	0.8
500 Msps/Cyclone V	225	192/150	0.485	3.5
1 Gsps/Kintex Ultrascale+	250	384/1300	0.210	~3(?)
1 Gsps/Kintex Ultrascale+	500	192/1300	0.210	~3(?)

Results:

Optimal clock frequencies and resulting power consumption for the first level trigger in respective environments:

Altera Cyclone V – 225 MHz, 3.5W
Kintex Ultrascale+ – 250 MHz, ~3W

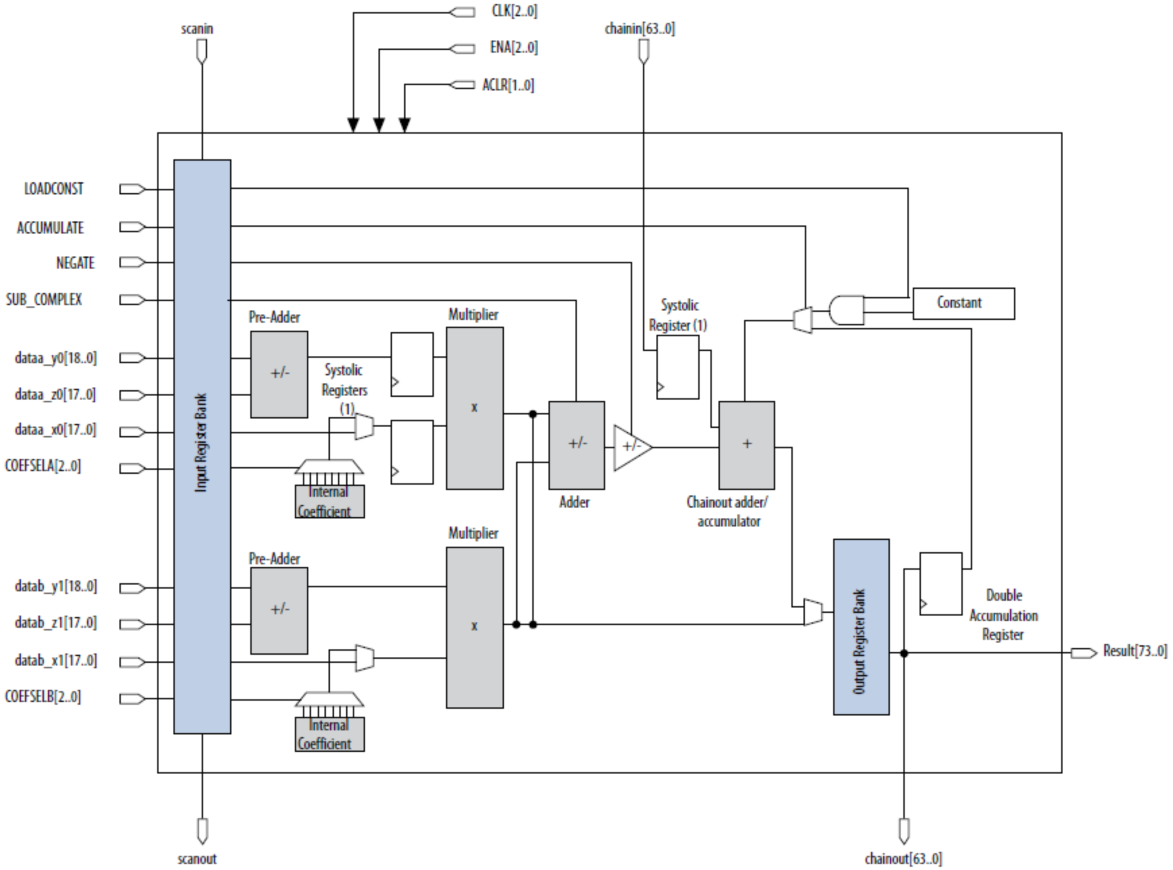
Conclusions:

Configurations suitable for:

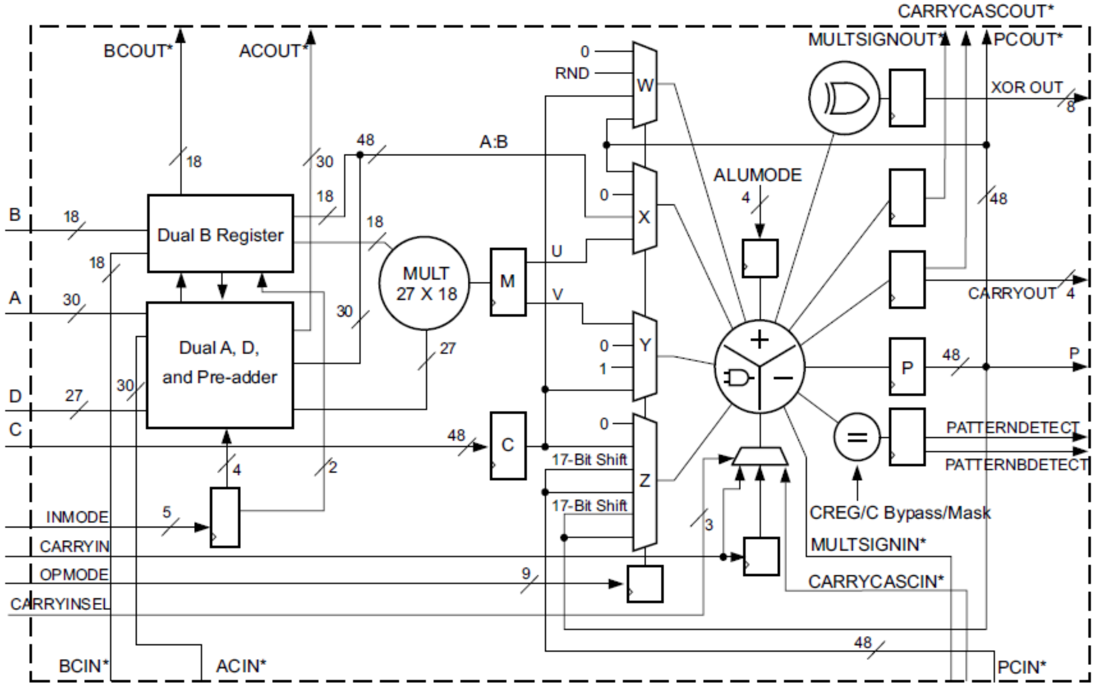
FLOWER second level trigger
FLOWER Theoretical first level trigger (doesn’t fit)
Future First level trigger (option 1)
Future First level trigger (option 1)

Multiplier internal structure - comparison

Cyclone V (28 nm)

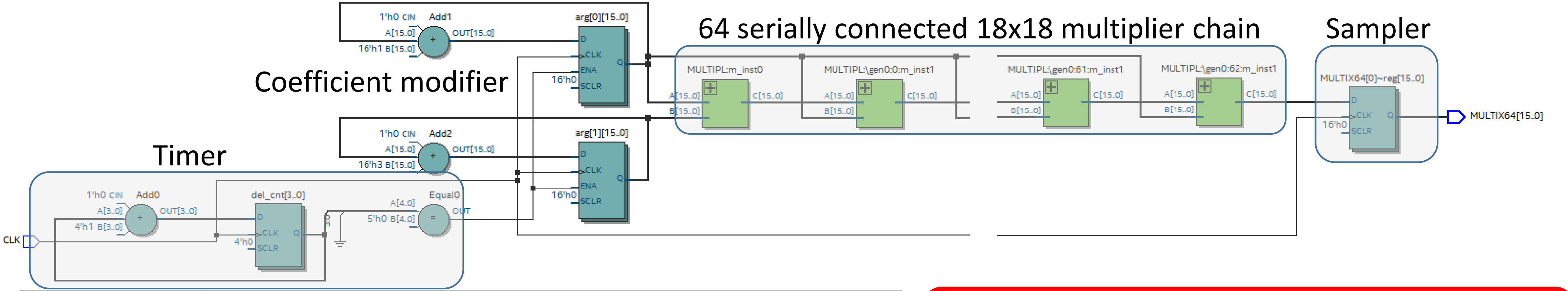


Ultrascale+ (16nm)



Multiplier performance

Performance check of the multipliers in the non-clocked chain: **Cyclone V**

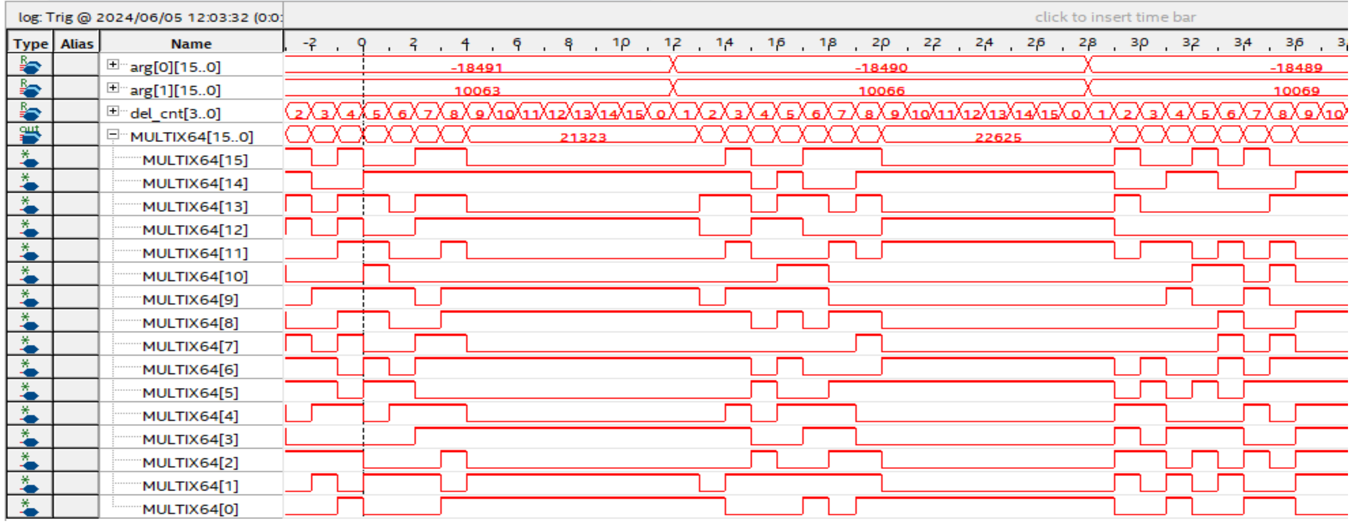


Coefficient modifier

64 serially connected 18x18 multiplier chain

Sampler

Timer



Results: Speed

The result stabilizes after 8 clock cycles (160 ns)

The average propagation time per multiplier

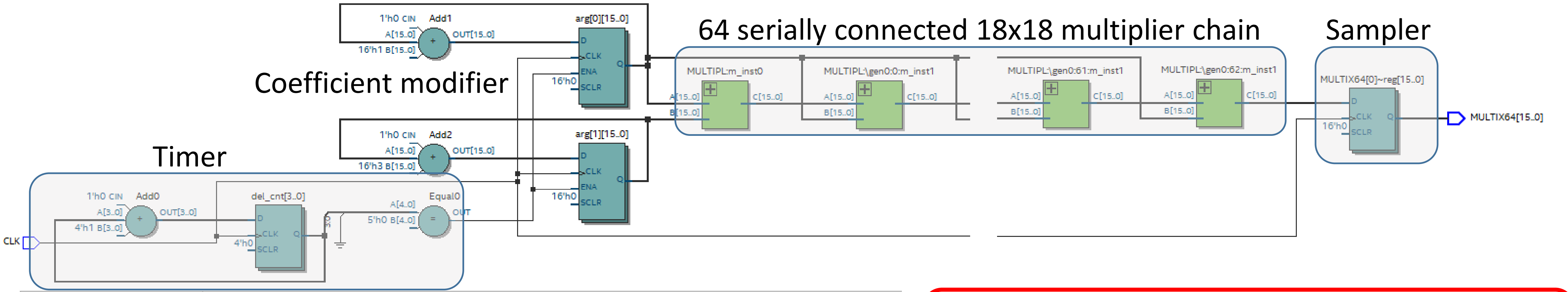
τ_m = 160ns/64 = **2.5 ns**

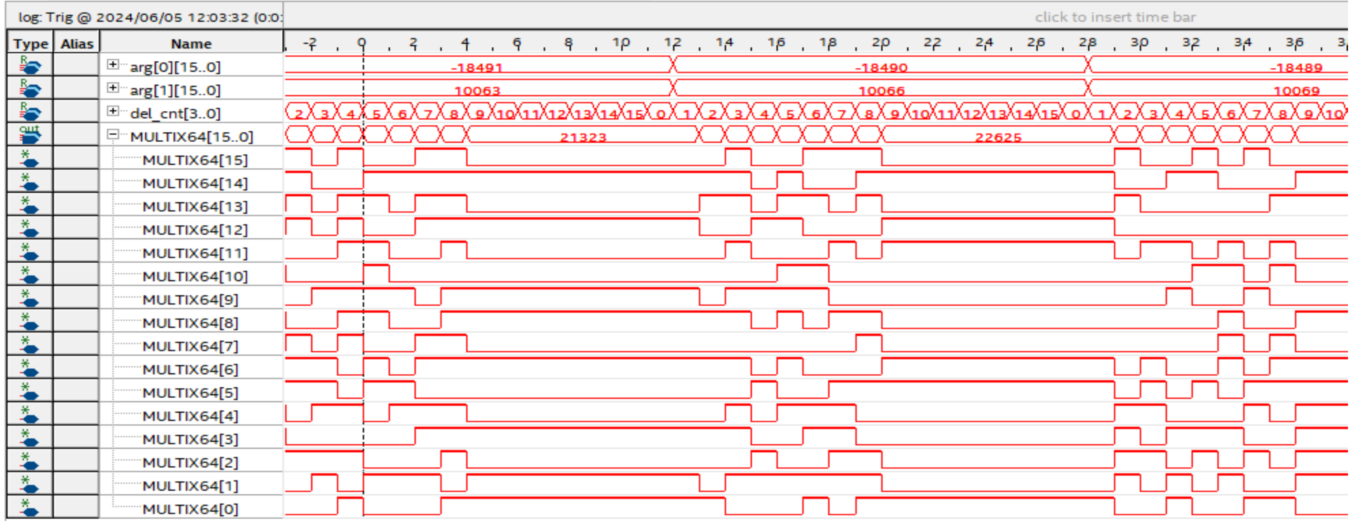
This is consistent with the device specification

f_{max} = 400 MHz

Multiplier performance

Performance check of the multipliers in the non-clocked chain: **Cyclone V**





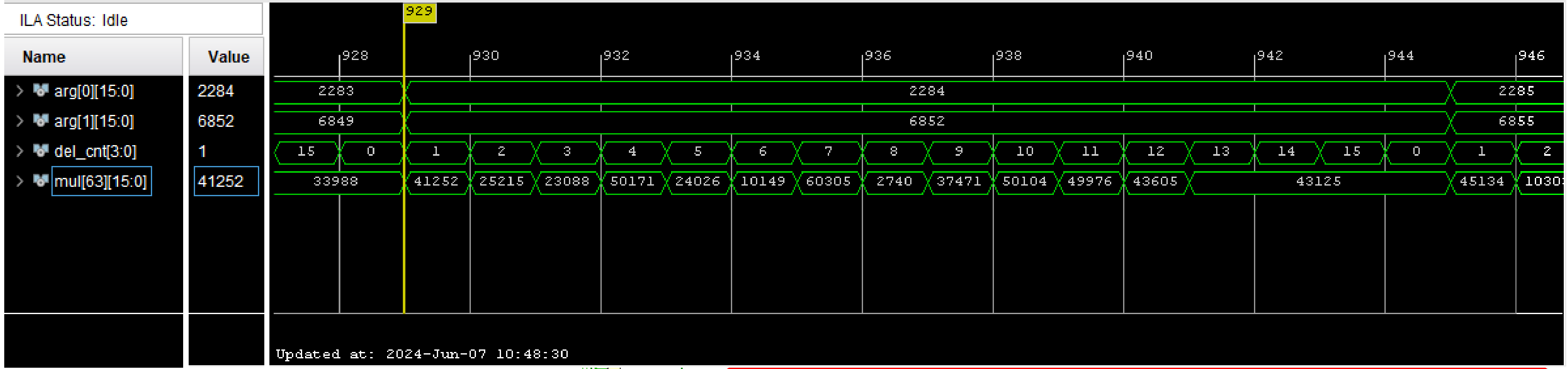
Results: Power consumption

The power consumption in this configuration amounts to 0.72W. Given ca 50% active time the average power consumption per multiplier can be estimated as:

P_M ***= 50uW/MHz*** per multiplier stage

Multiplier performance

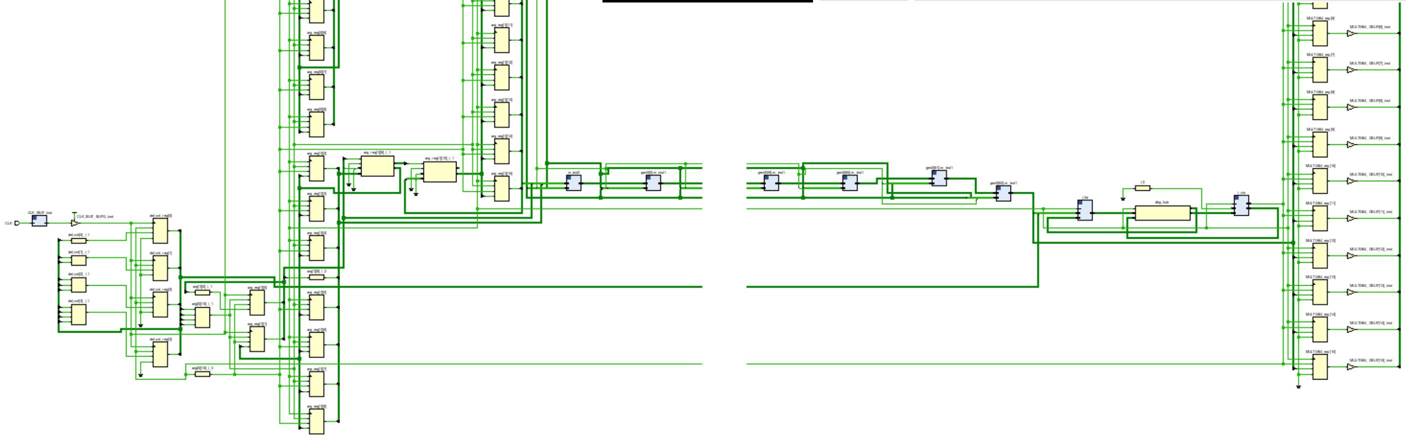
Performance check using 64 multipliers in the non-clocked chain: **Kintex Ultrascale+**



ILA Status: Idle

Name	Value
arg[0][15:0]	2284
arg[1][15:0]	6852
del_cnt[3:0]	1
mul[63][15:0]	41252

```
           929
      928       930     932     934     936     938     940     942     944     946
   2283   |                          2284                                   |  2285
   6849   |                          6852                                   |  6855
 15 |  0  |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |  9  | 10  | 11  | 12  | 13  | 14  | 15  |  0  |  1  |  2
  33988   |41252|25215|23088|50171|24026|10149|60305|2740 |37471|50104|49976|43605|        43125        |45134|1030
Updated at: 2024-Jun-07 10:48:30
```



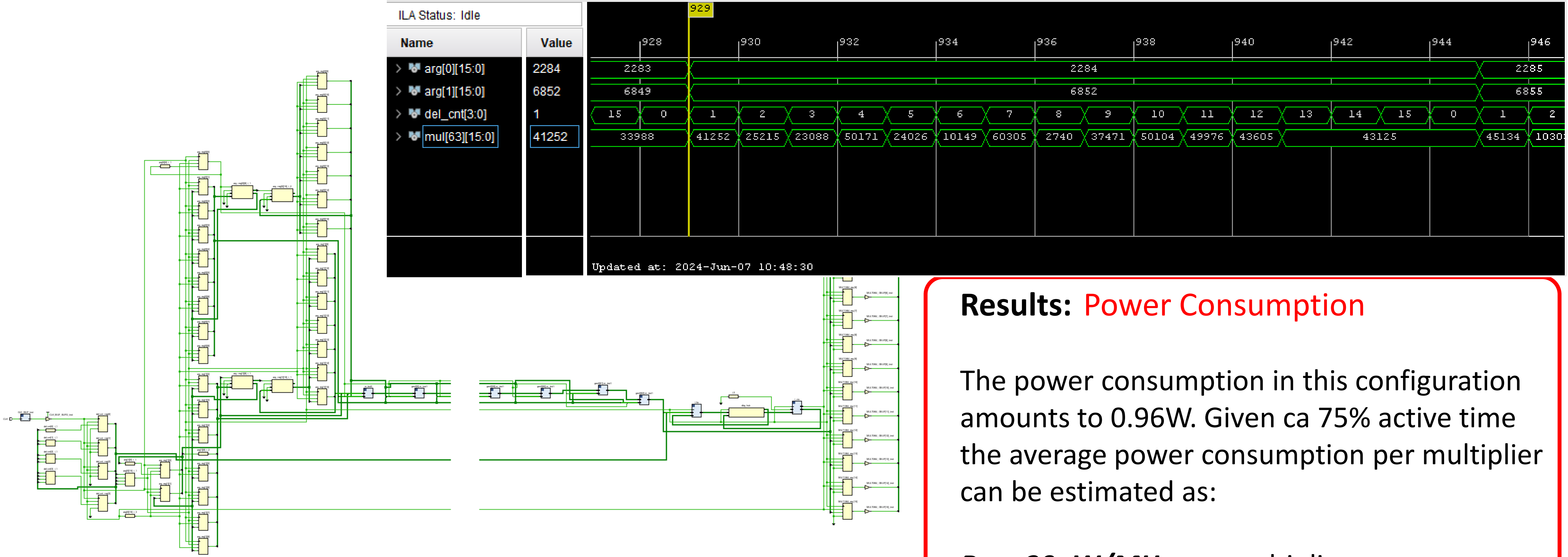
Results: Speed

The result stabilizes after 12 clock cycles (60 ns)
The average propagation time per multiplier
τ_m = 60ns/64 = **0.94 ns**

This is consistent with the device specification
f_{max} > 1 GHz

Multiplier performance

Performance check using 64 multipliers in the non-clocked chain: **Kintex Ultrascale+**



ILA Status: Idle

Name	Value
arg[0][15:0]	2284
arg[1][15:0]	6852
del_cnt[3:0]	1
mul[63][15:0]	41252

Updated at: 2024-Jun-07 10:48:30

Results: Power Consumption

The power consumption in this configuration amounts to 0.96W. Given ca 75% active time the average power consumption per multiplier can be estimated as:

P_M ***= 20uW/MHz*** per multiplier stage

Conclusions

1. The first version of the second level Convolutional Neural Network Trigger for the RNO-G and IceCube-Gen2 has been constructed and verified using pre-loaded data records.
2. We have performed feasibility studies for the continous CNN trigger regarding achievable clocking frequencies, firmware architectures and power consumption.
3. For the tests we have used an Intel DE1-SoC evaluation board featuring a Cyclone V FPGA and a custom-made PANDA-DC prototype of the Data Concentrator for the PANDA experiment
4. The design is practically entirely written in a platform independent VHDL.

Thank You!