ONLINE MACHINE LEARNING BASED EVENT SELECTION FOR COMET PHASE-I

Yuki Fujii (Monash University),
M. Miyataki, Y. Nakazawa, L. Pinchbeck, H. Yoshida, K. Ueno, M. J. Lee

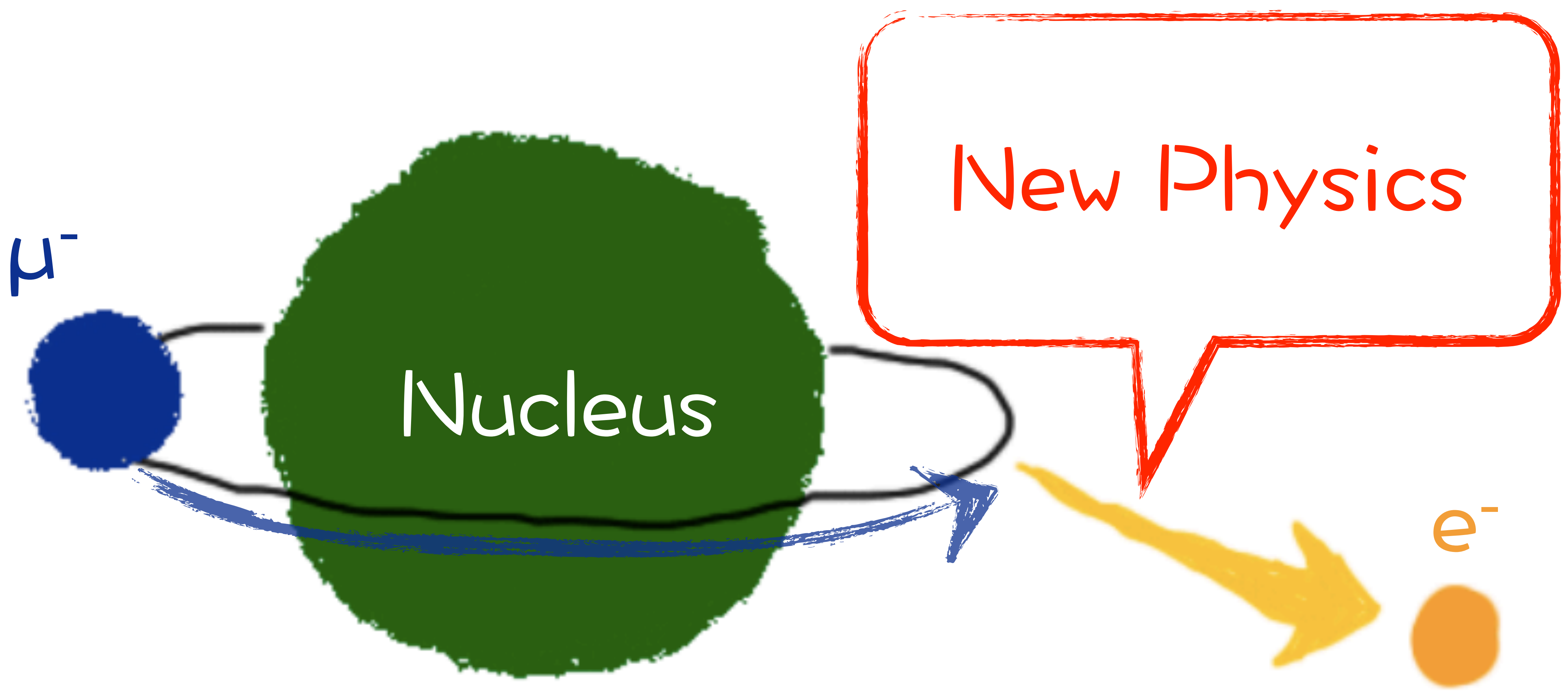
The 23rd International Workshop on Neutrinos from Accelerators
5th August 2022, Salt-Lake city, Utah

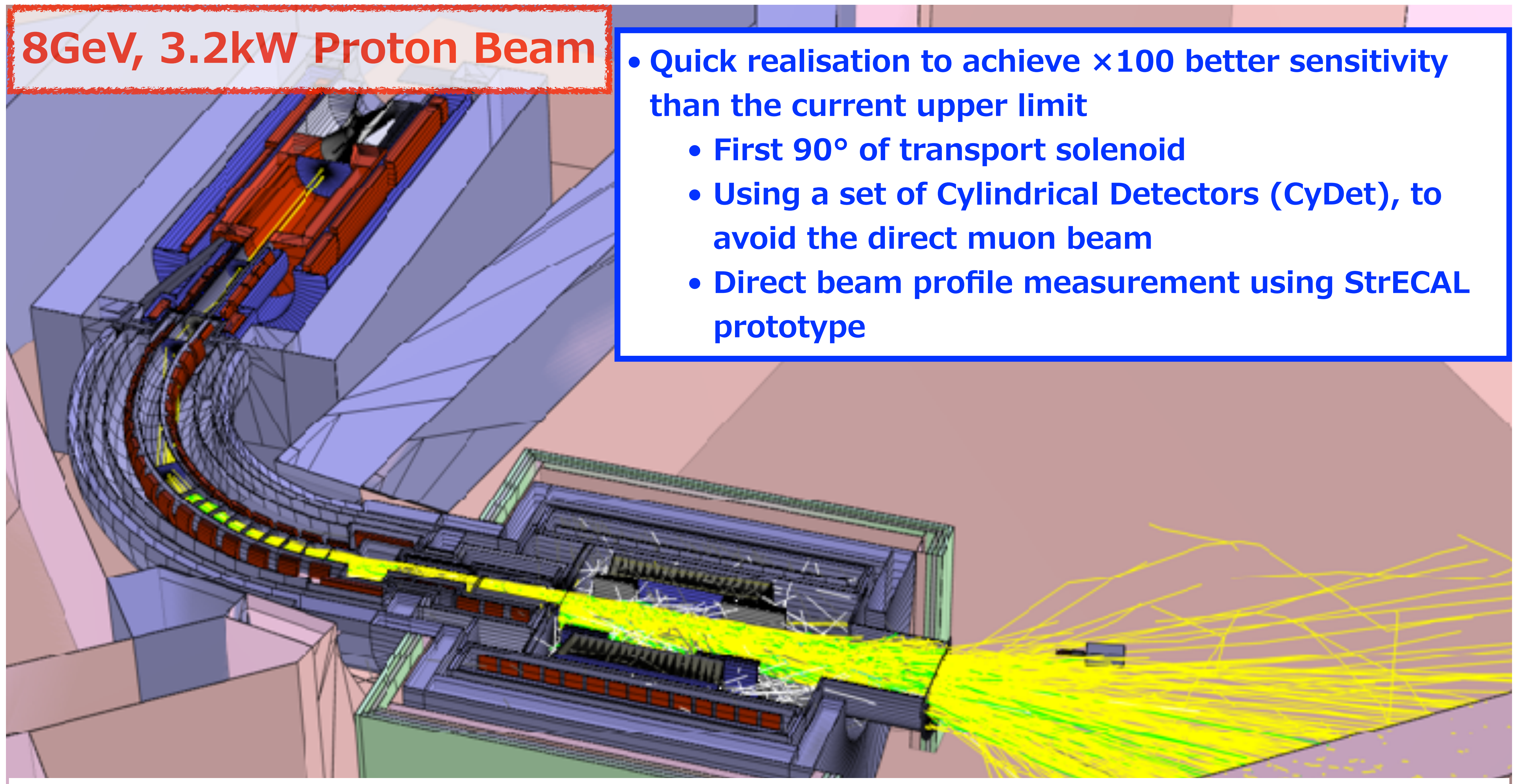
INTRODUCTION

- Everyone wants a fast, high efficiency and high purity trigger
  - +cost-effectiveness, redundancy, quick realisations etc…
  - In general, it directly determines the signal efficiency (=experimental sensitivity) of our experiments
  - **We want more data, more physics, more, more & more…**
  - Several solutions
    - Trigger-less (offline trigger w/ GPUs), hardware level vetos, extremely fast data pipeline + gigantic data storage…
    - … or using field programmable gate arrays (FPGAs)
  - **This talk is based on arXiv:2010.16203 (Y. Nakazawa *et.al.*) + new studies mainly done by M. Miyataki**

COMET EXPERIMENT PHASE-I

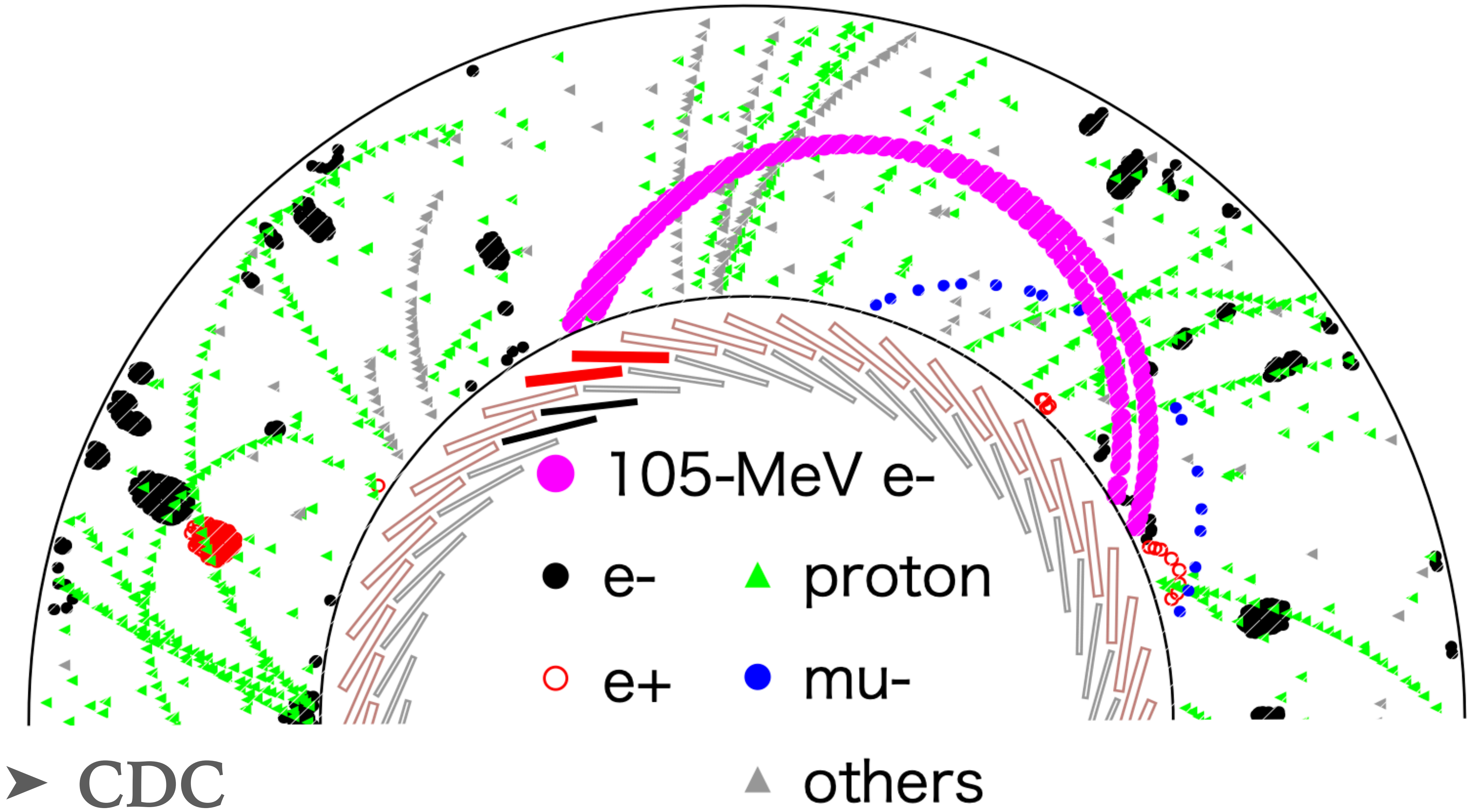
- Searching for a **μ-e conversion** with sensitivity of $\boldsymbol{O(10^{-15})}$ in its Phase-I
  - Muon beam produced by impinging the 8 GeV proton beam onto the graphite target
  - Requires $\sim 10^{18}$ total stopping muons per 150 days → $\mathbf{10^{10}}$ $\boldsymbol{\mu^{-}}$**/sec**
  - So many secondary particles will be expected inside the detectors
- **See Sam Dekkers talk for more details**

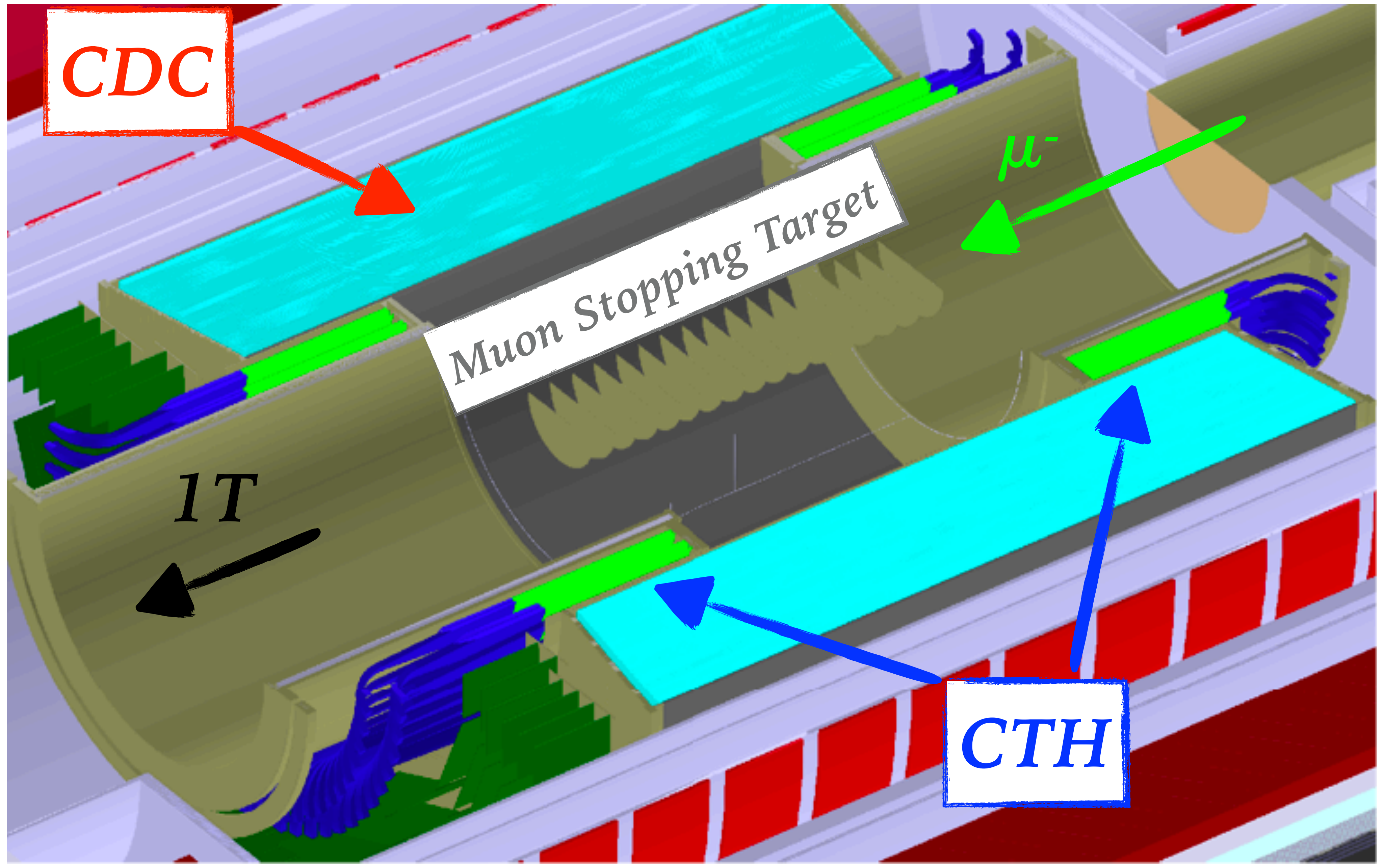




COMET Phase-I technical design report, *PTEP*, Vol 2020, Issue 3, March 2020, 033C01, https://doi.org/10.1093/ptep/ptz125

CYLINDRICAL DETECTOR (CYDET)





- **CDC** C. Wu, *et.al.* DOI:10.1016/j.nima.2021.165756
  - ~5,000 wires, 20 full-stereo layers for momentum measurement, typical drift time <400ns
  - Signal electrons' trajectories fully contained inside the volume
- **CTH** Y. Fujii, *et.al.* DOI:10.5281/zenodo.6781368
  - 2 layers of 64 segmented plastic scintillator rings at both ends of CDC for the timing measurement
  - Suppress accidental events and low momentum particles by taking four-fold coincidence

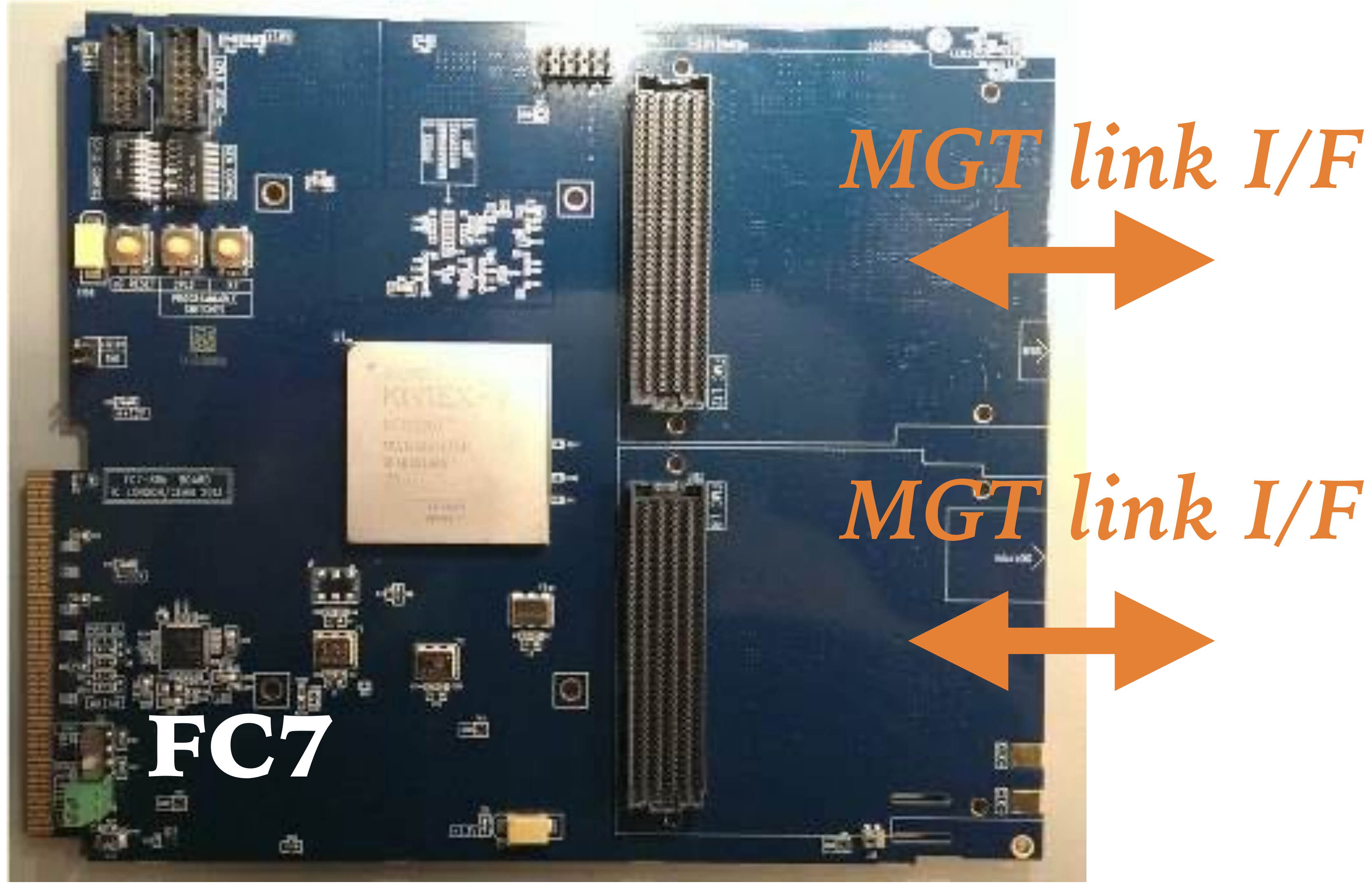
TRIGGER REQUIREMENTS

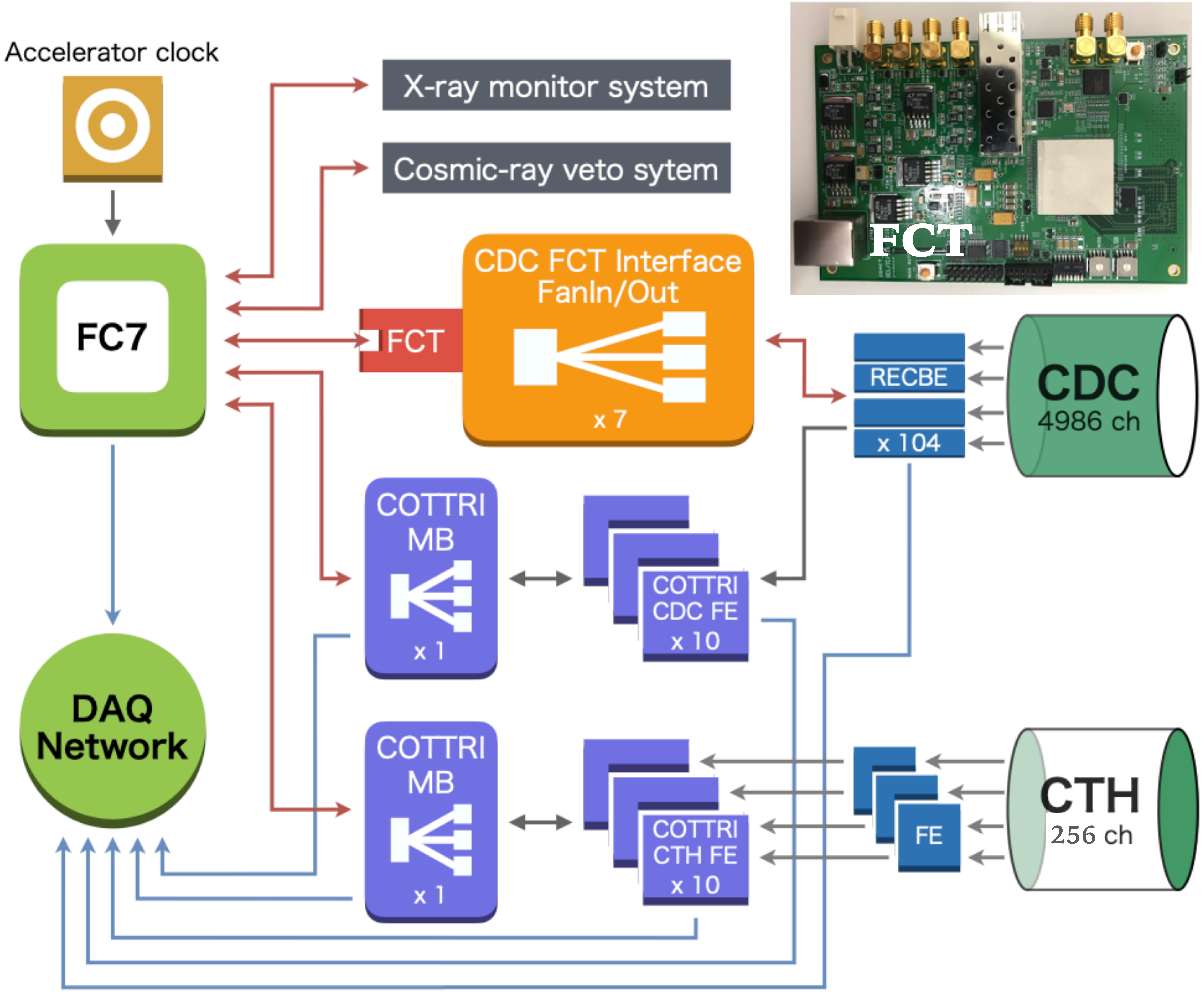
- **Strong fake trigger suppression**
  - Expected 4 fold coincidence rate is **~90kHz** from fake events in CTH

    ⟺ DAQ system requires **<13kHz** trigger rate (bottleneck = data processing rate)
  - At least **1/7** further suppression is needed while keeping the high signal acceptance
- **Fast online event selection**
  - Less than **7 μsec** latency is allowed (limited by the online buffer size)
- **Flexibility**
  - Availability of the timely modification for possible changes in situations (BG rate, etc)
  - Multiple triggers (bi-products, calibrations, BG enriched etc.)
- **Stability**

COMET CENTRAL TRIGGER SYSTEM IN PHASE-I

- FC7 + FCT
  - Make a final trigger decision based on CDC trigger info + CTH trigger info + accelerator info
  - Distribute the trigger signal & a 40MHz common clock to all readout and trigger modules

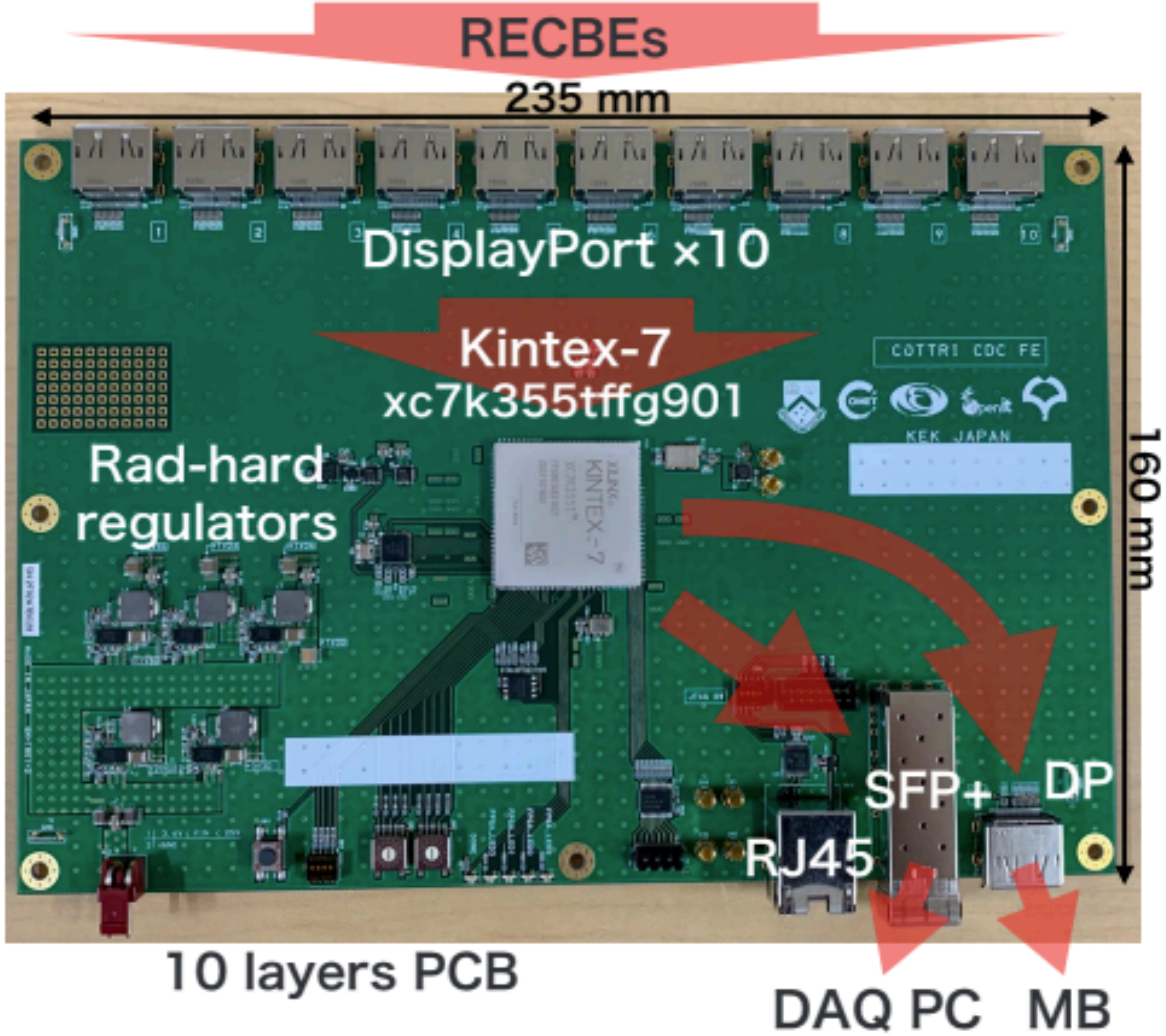




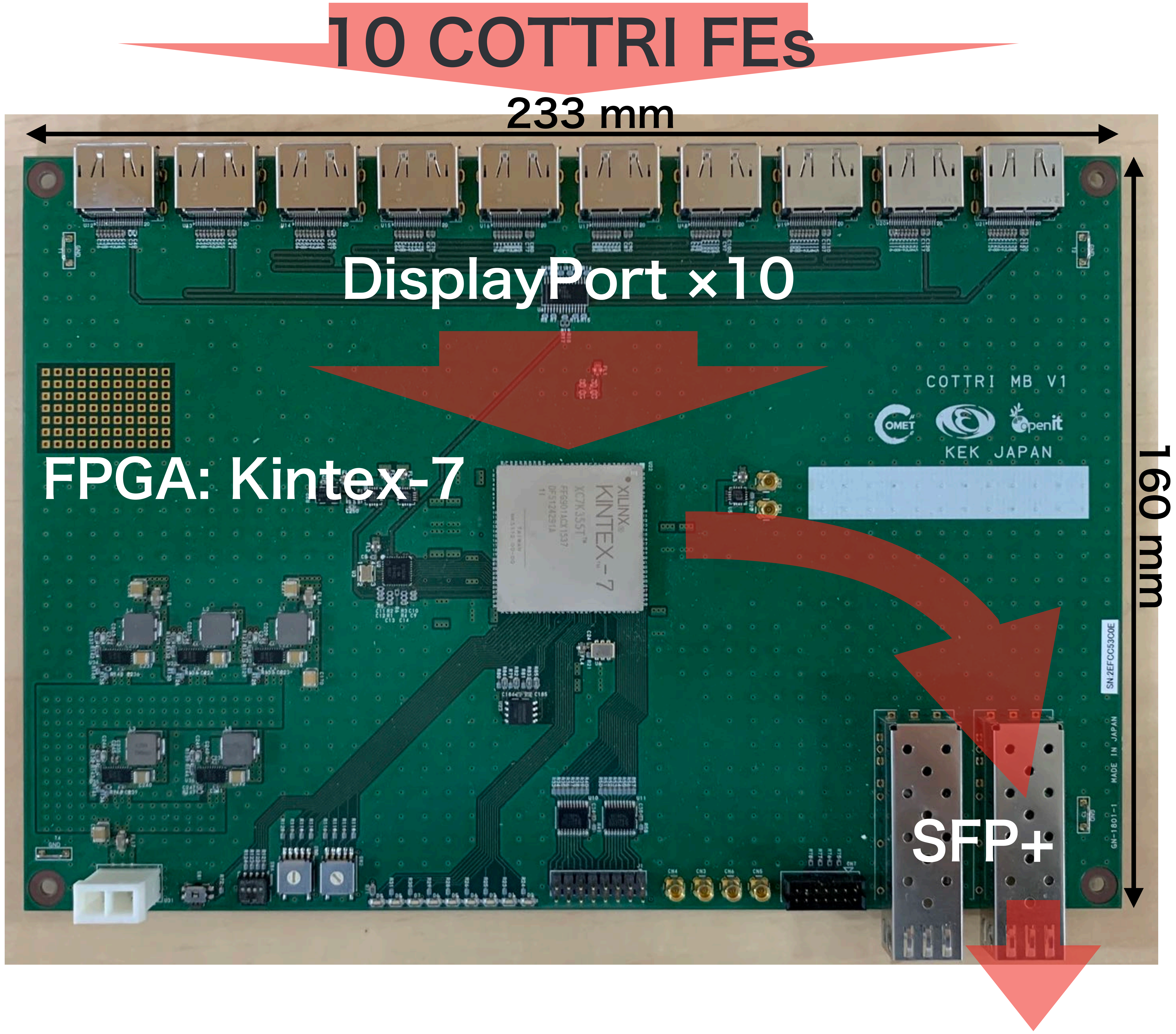
COTTRI SYSTEM (1)

COTTRI SYSTEM (2)

- **COTTRI CDC FE**
  - Purely digital processing board by utilising FPGA (Kintex-7) and Multi-Gibabit data Transfer technologies (MGT link)
  - 10 boards cover 100 CDC readout boards corresponding to 4,800 wires
  - Perform hit classifications to identify more signal-like hits compared to other proton/low-e hits
  - Send those information to COTTRI merger board through MGT link



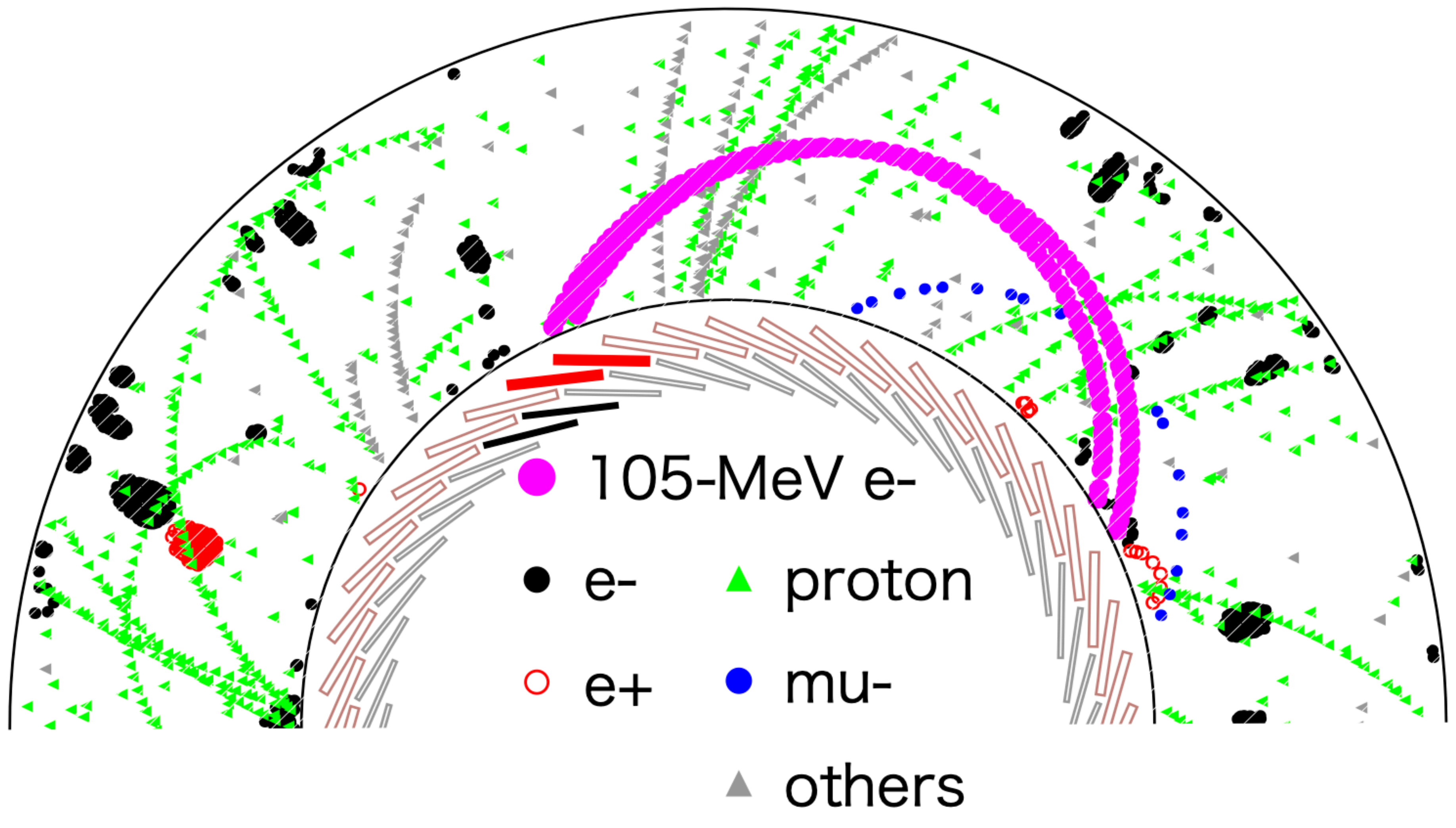
COTTRI SYSTEM (3)

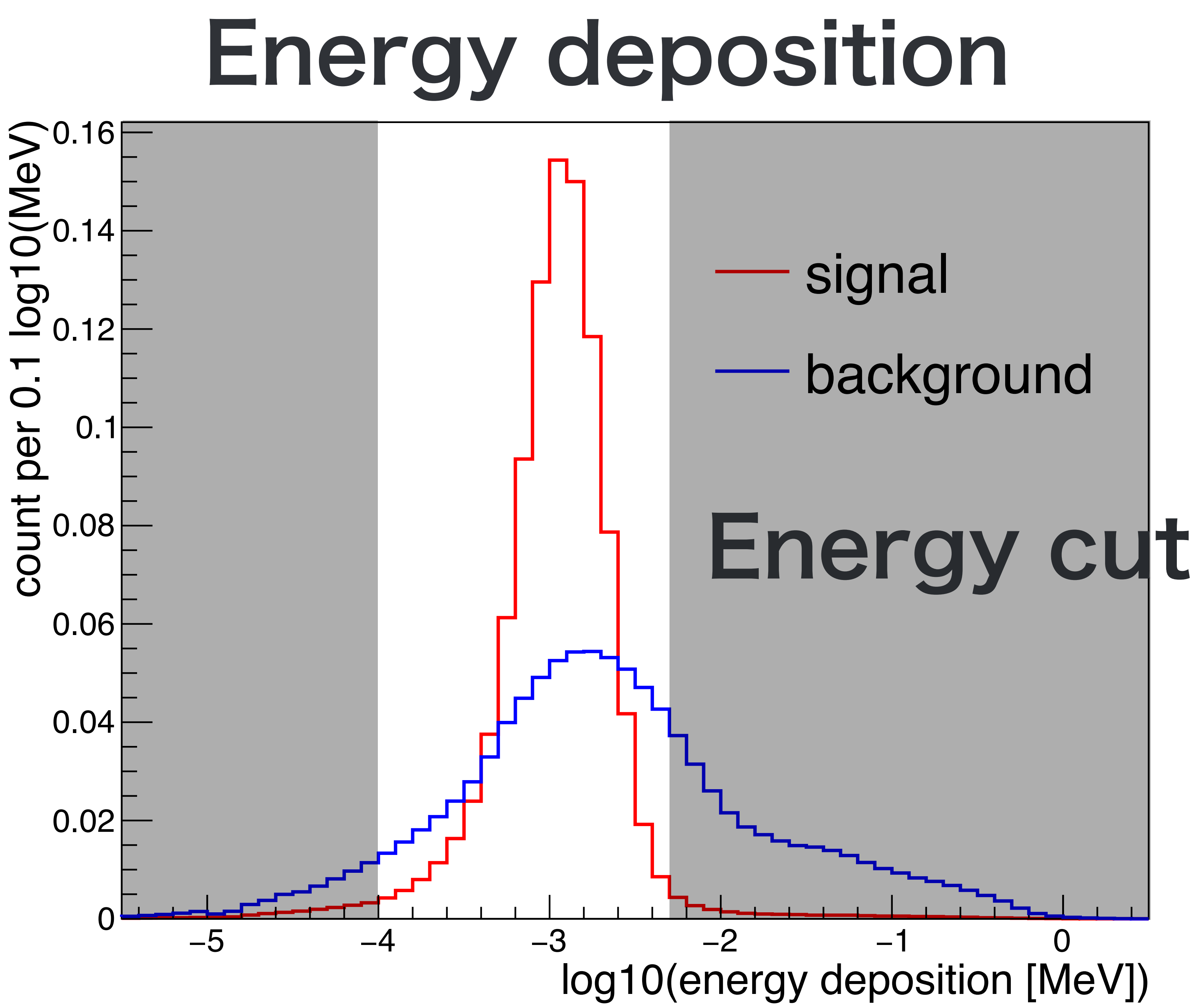


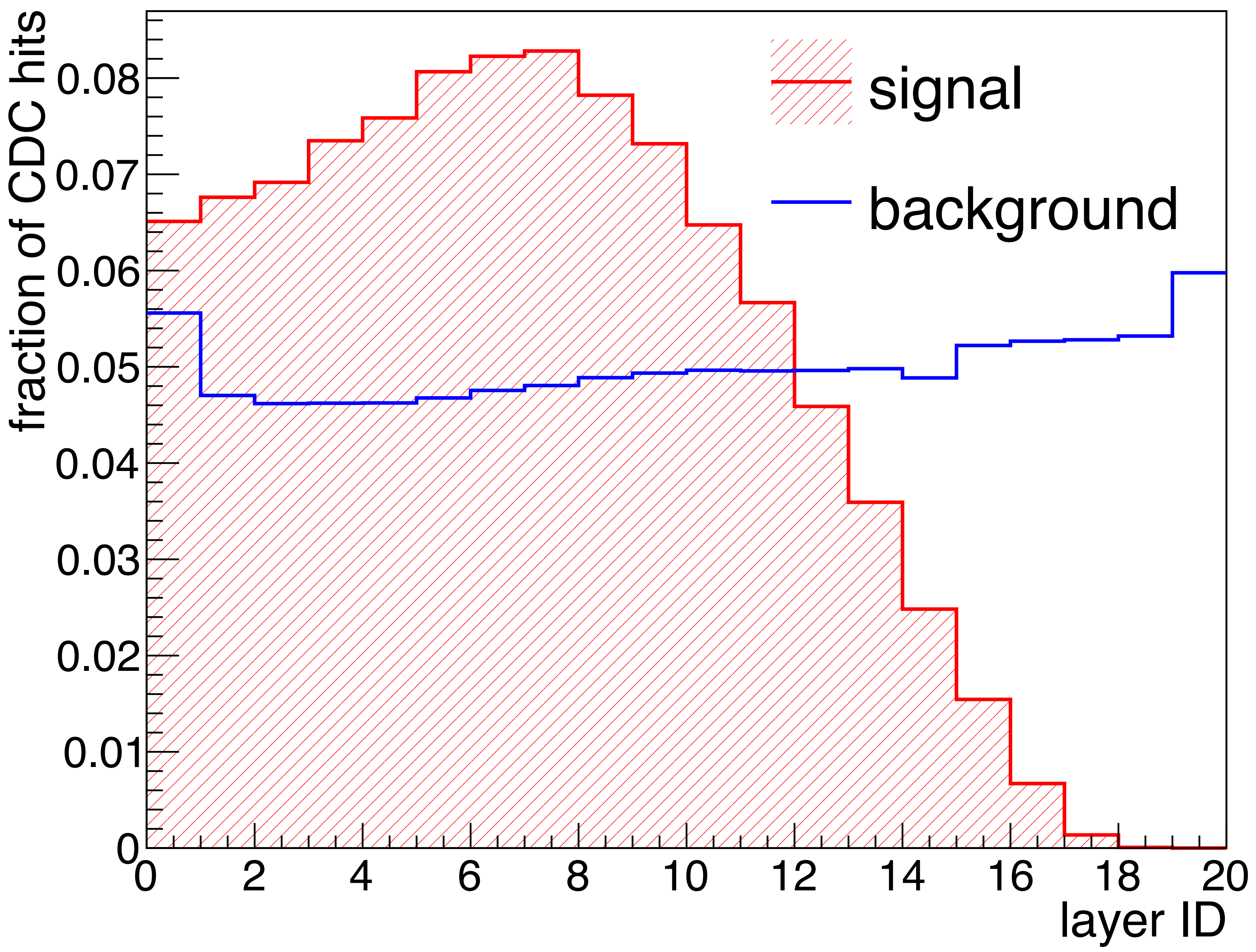
- **COTTRI CDC MB**
  - Similar to COTTRI CDC FE (same FPGA, same MGT links)
  - Cover all 10 COTTRI CDC FE boards with one MB
  - Perform event classifications based on the hit information
  - Send final CDC trigger info to the central trigger system (FC7) via MGT link

DECISION TREE BASED HIT CLASSIFICATION (1)

- Separate signal/background-like hits by using Gradient Boosted Decision Tree (GBDT)
- BG hits mostly induced by protons & low momentum e- (from γ/n)
  - Larger dE/dx, uniform layer distributions, less neighbouring hits
  - Training of GBDT model can be done offline by using MC/real data
    - The tables with weights generated and to be implemented inside the FPGA



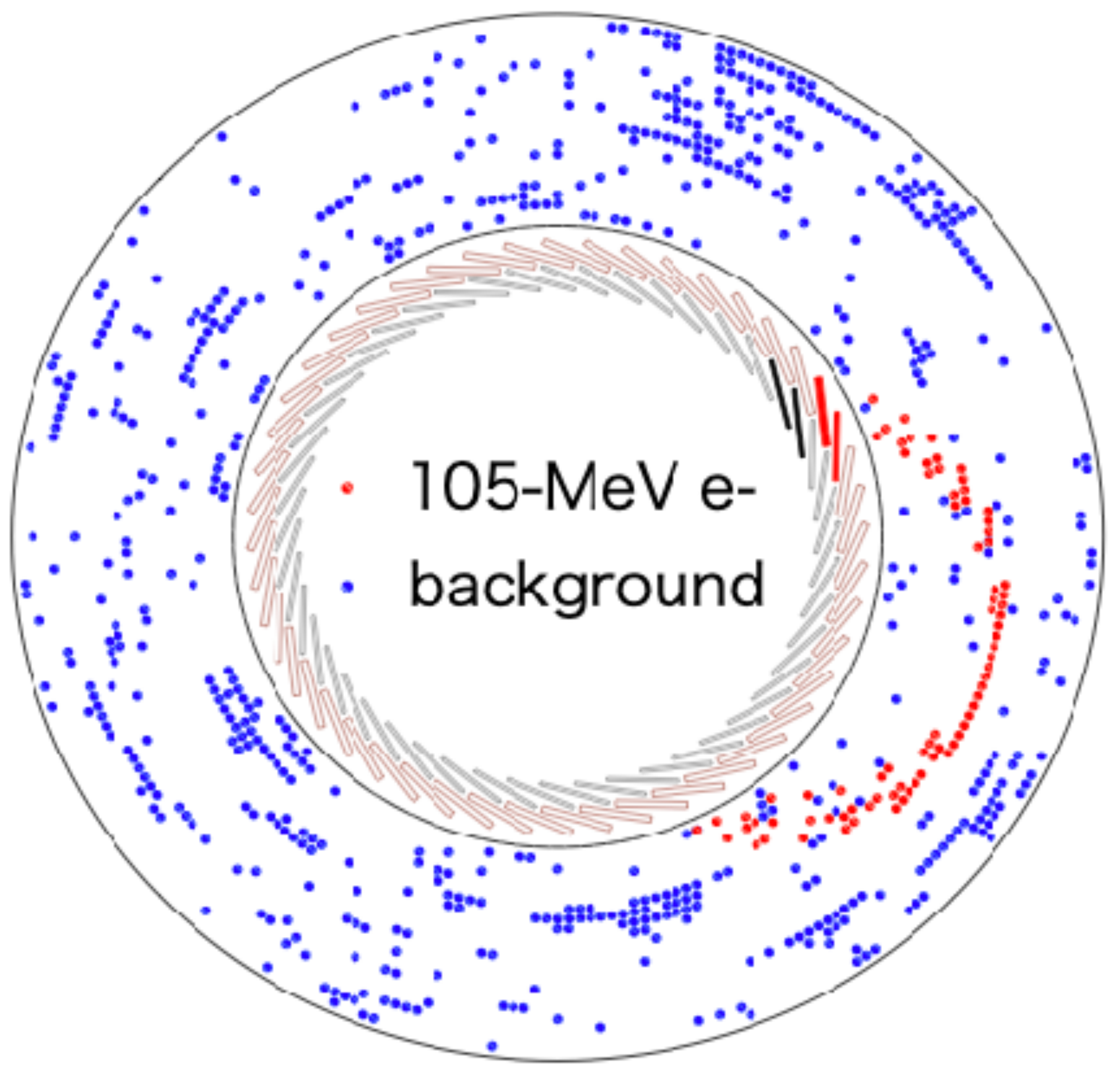


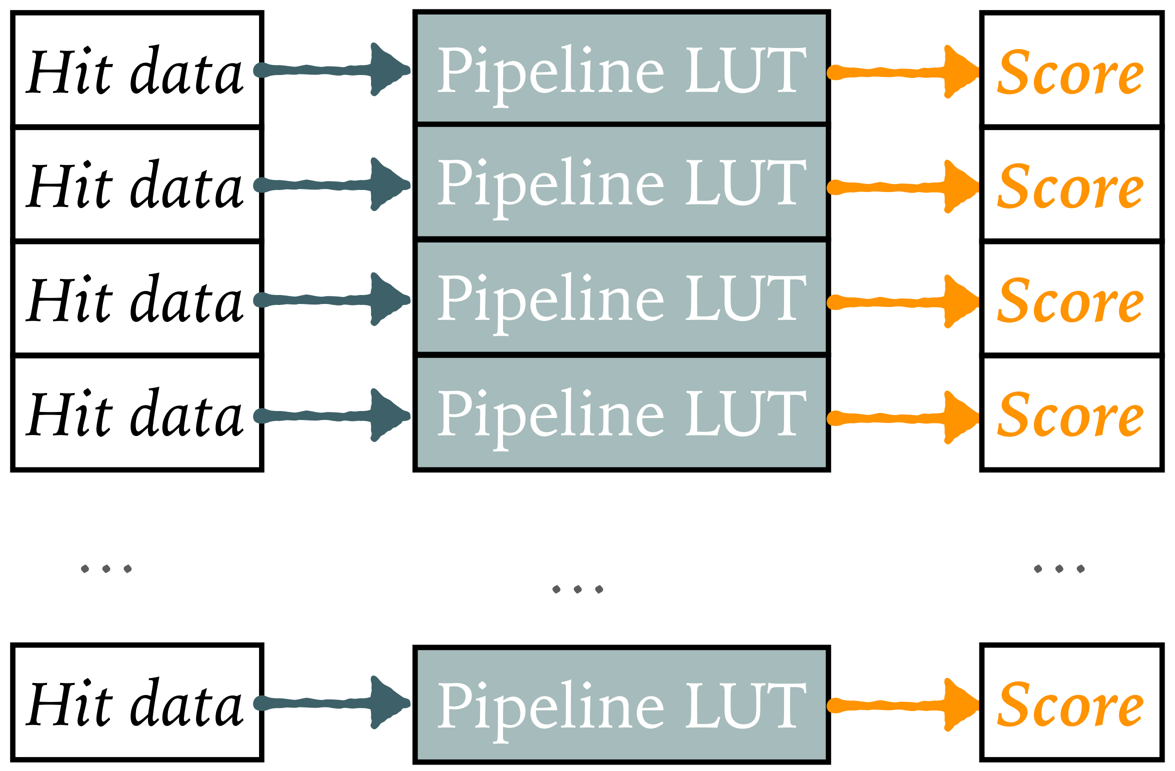


DECISION TREE BASED HIT CLASSIFICATION (2)

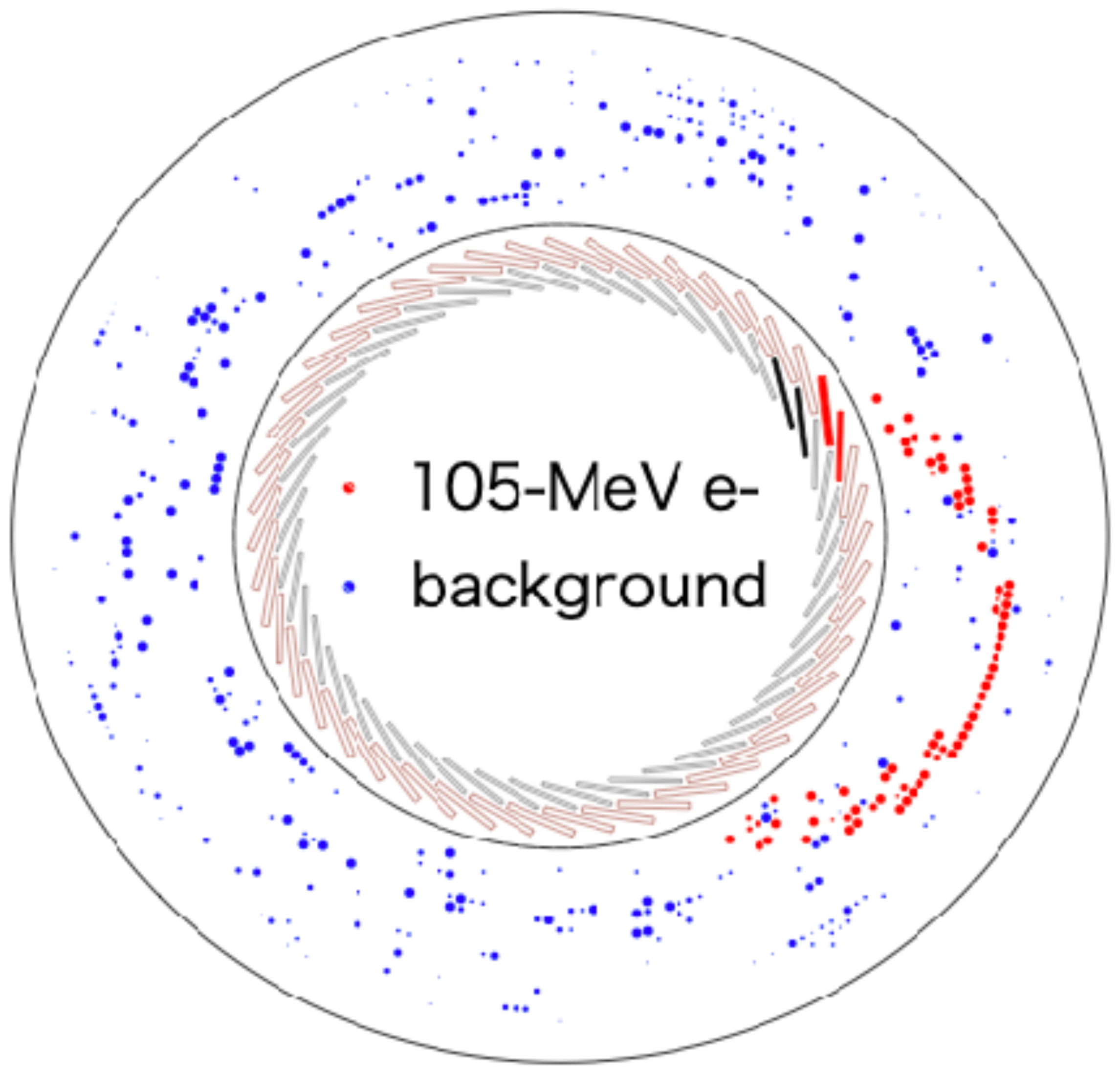
- Actual implementation
  - Perform hit classification by configuring look-up tables (LUTs) with GBDT weighting tables
  - One COTTRI CDC FE covers 10 RECBEs = 480 wires, 6-bit (2-bit ADC+neighbouring ADCs) data/each as input, decision tree's score as 6-bit output (larger = signal-like)
    - Only one or two clock cycles for the score calculation

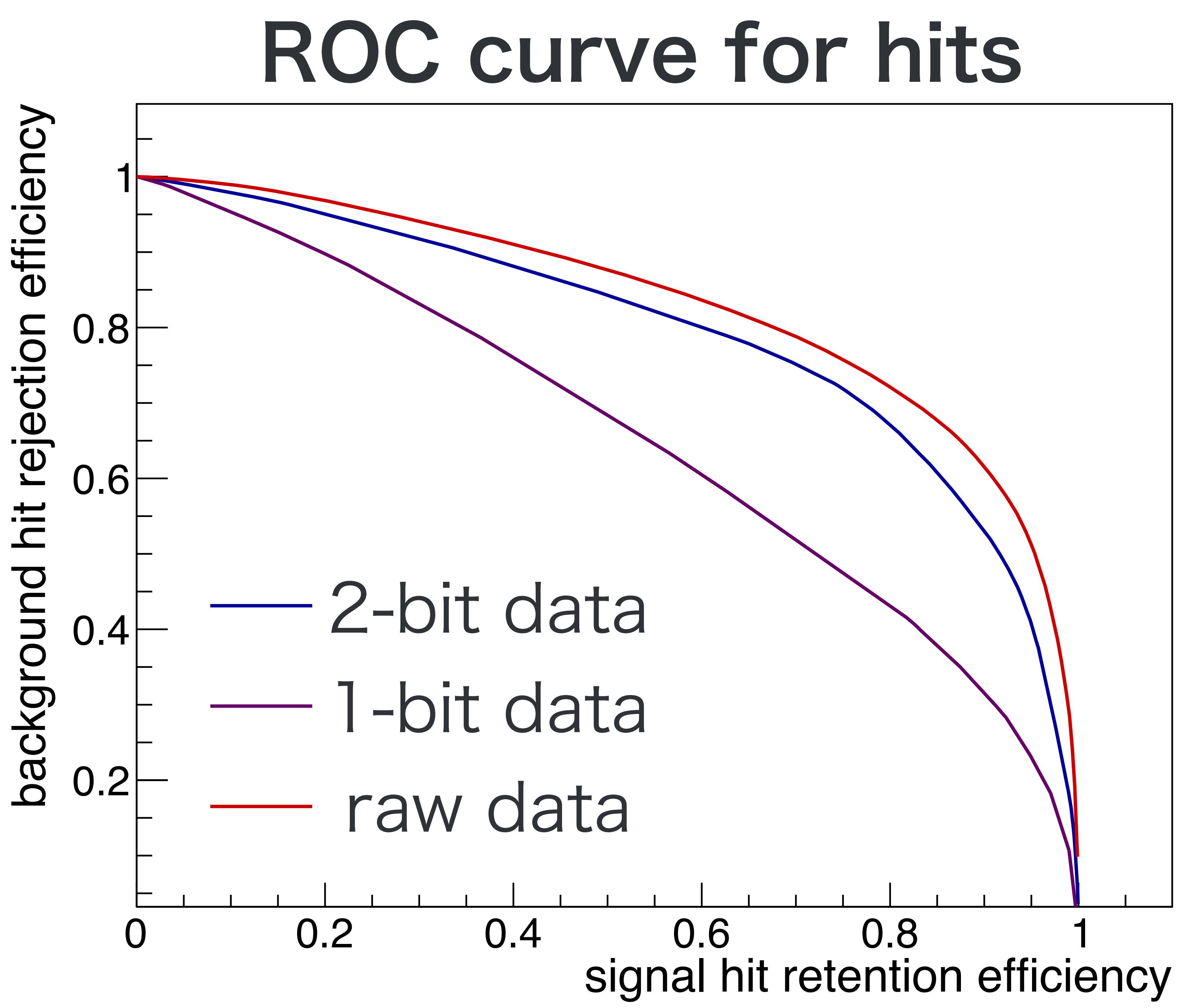
All projected hits in a single time window





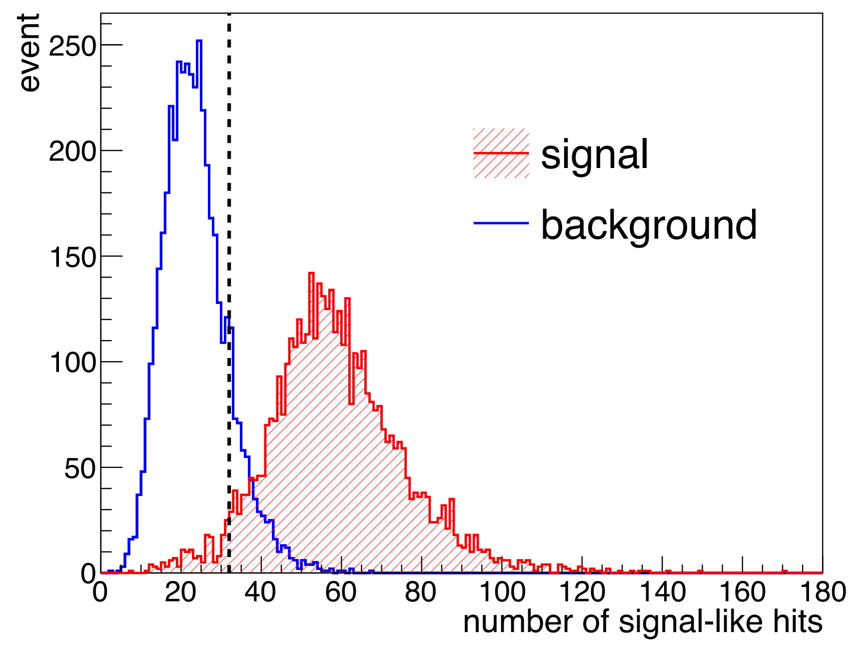
After scoring hits

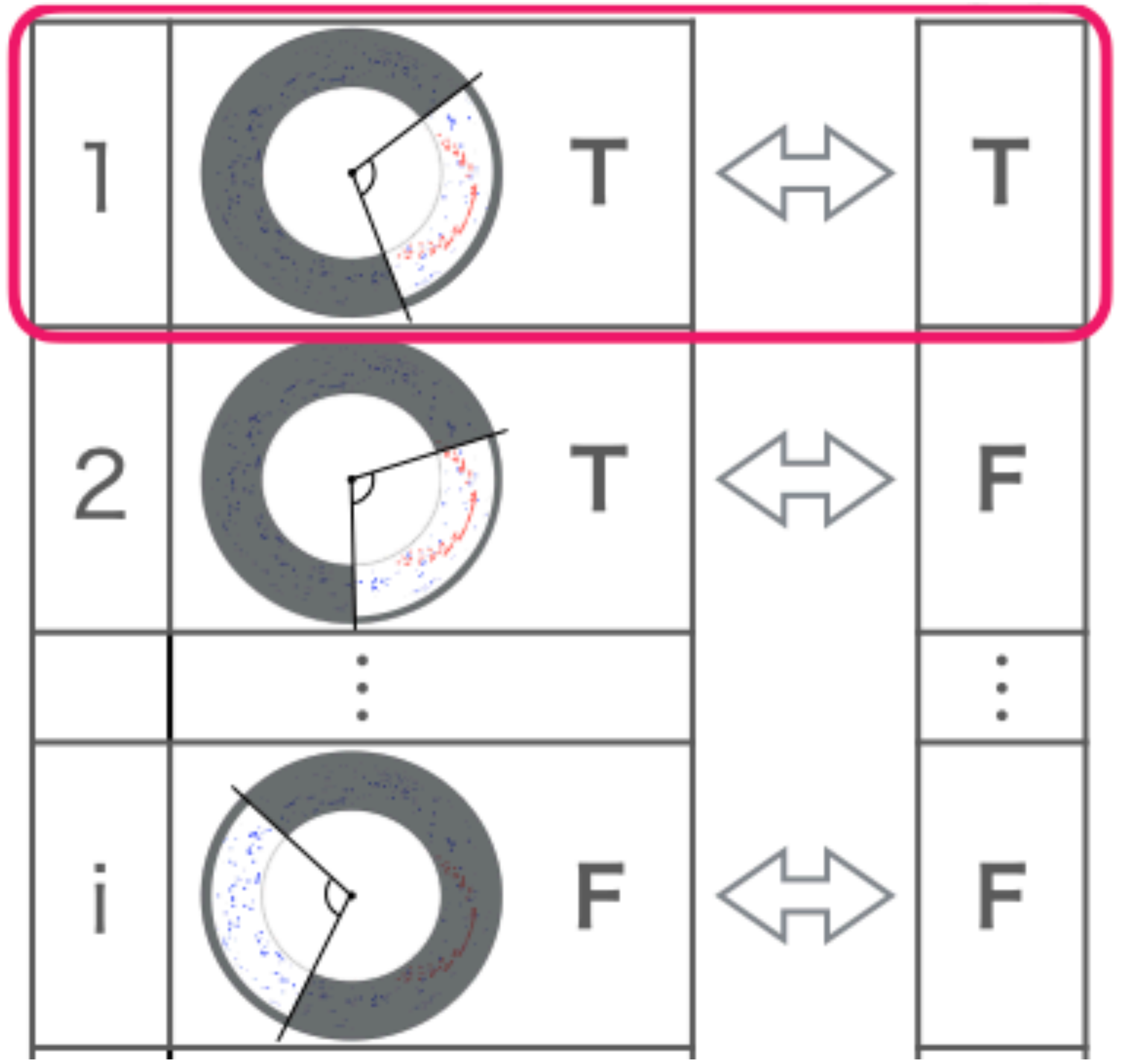


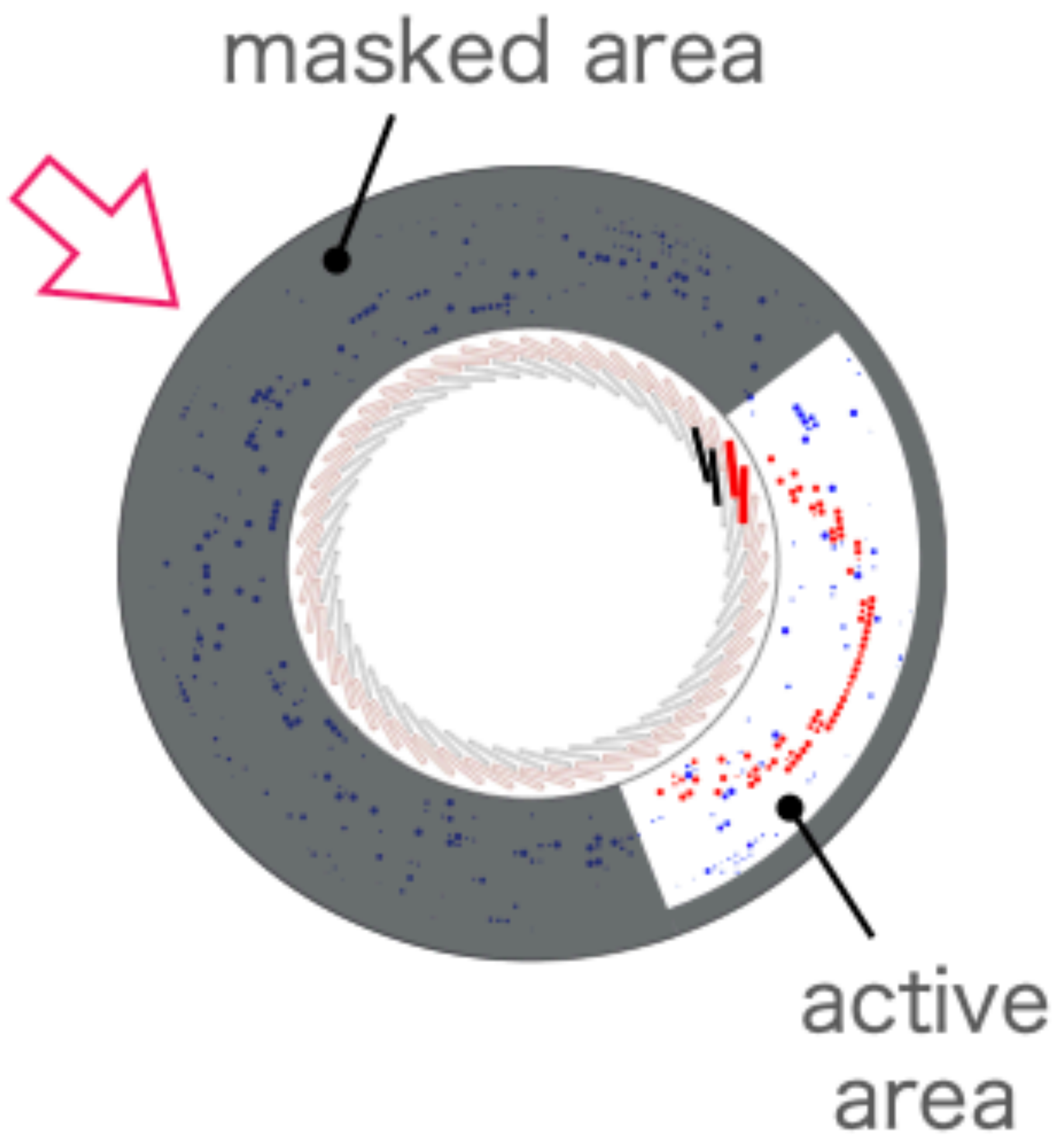


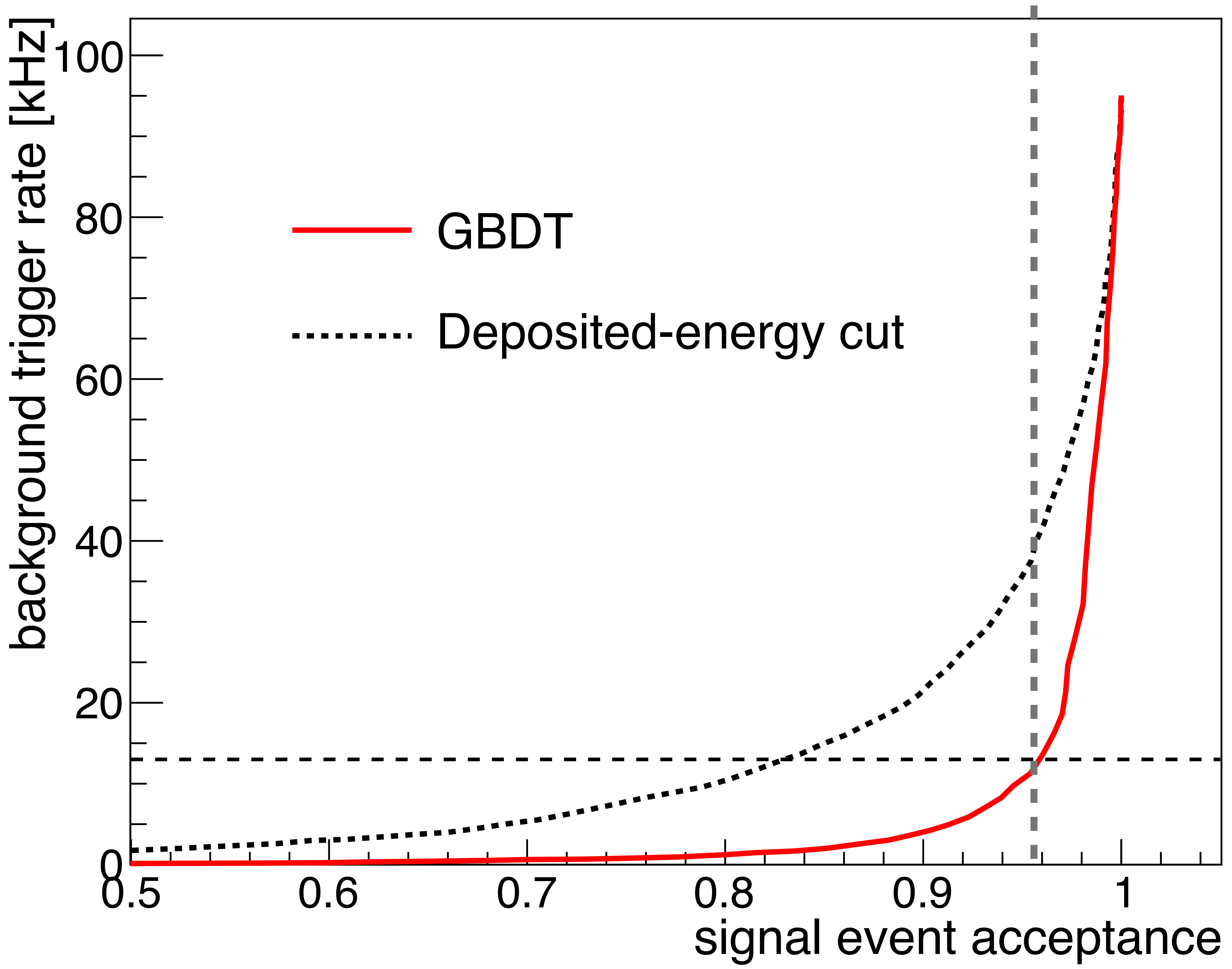
DECISION TREE BASED HIT CLASSIFICATION (3)

- Apply the geometrical cut to select the region of interest
- **96% signal event trigger acceptance @13kHz** (simulation)
  - **×3** stronger BG suppression compared to the cut based hit selection





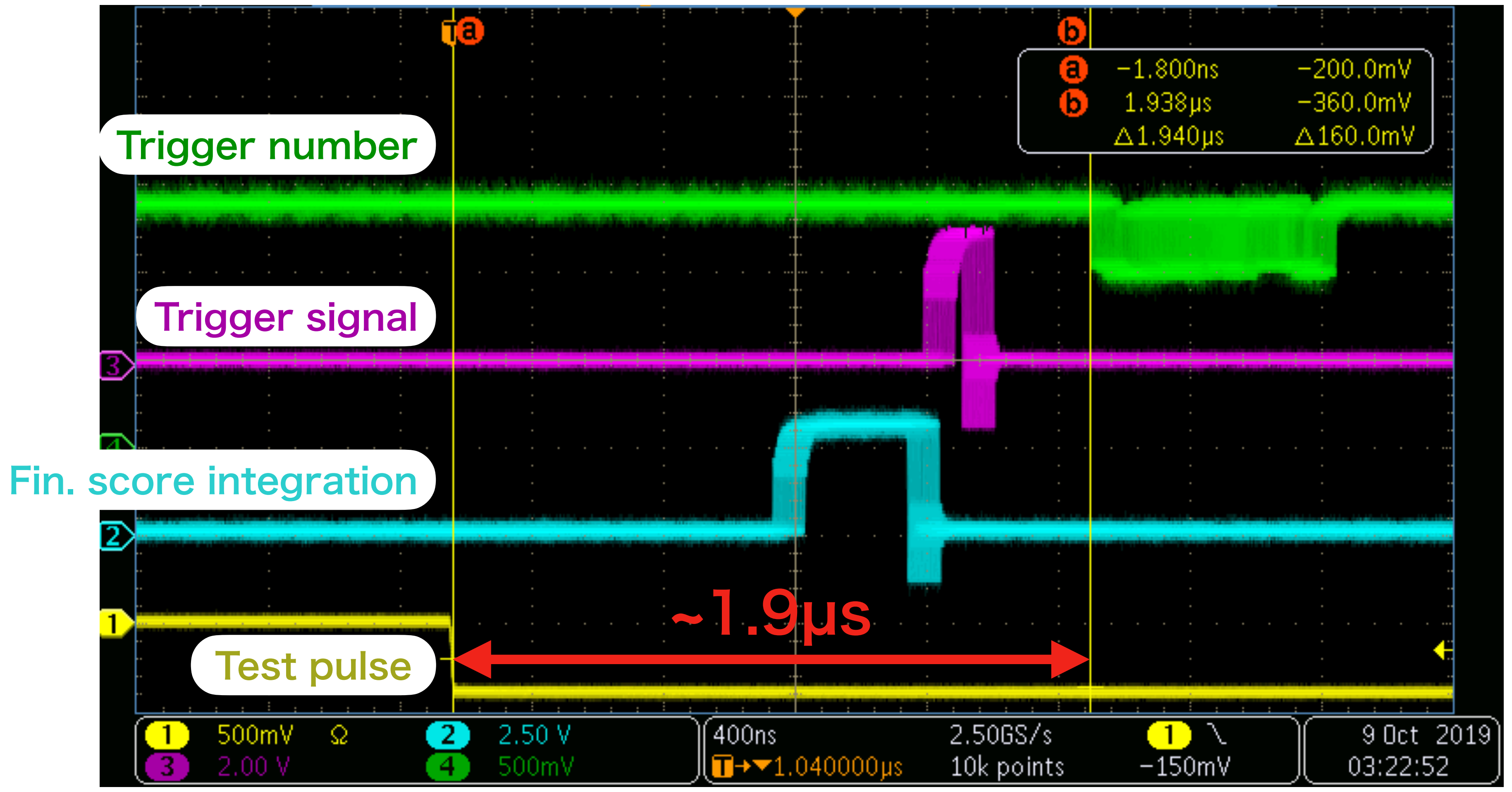


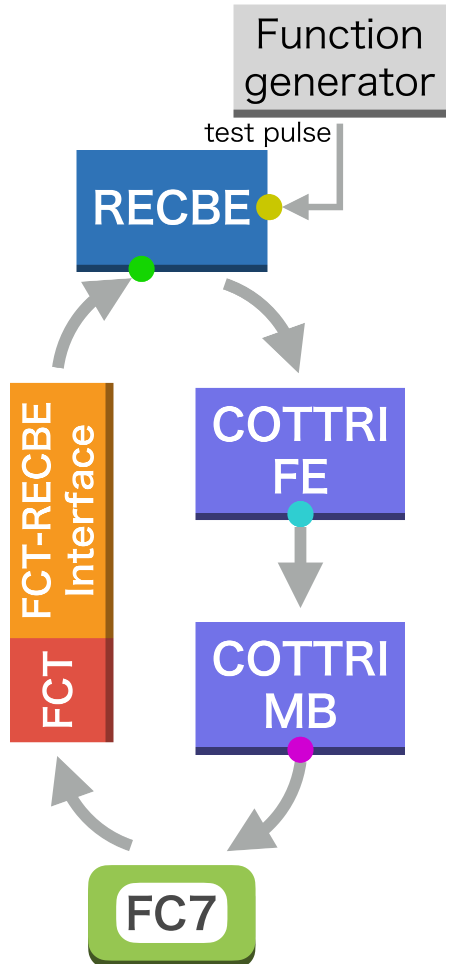


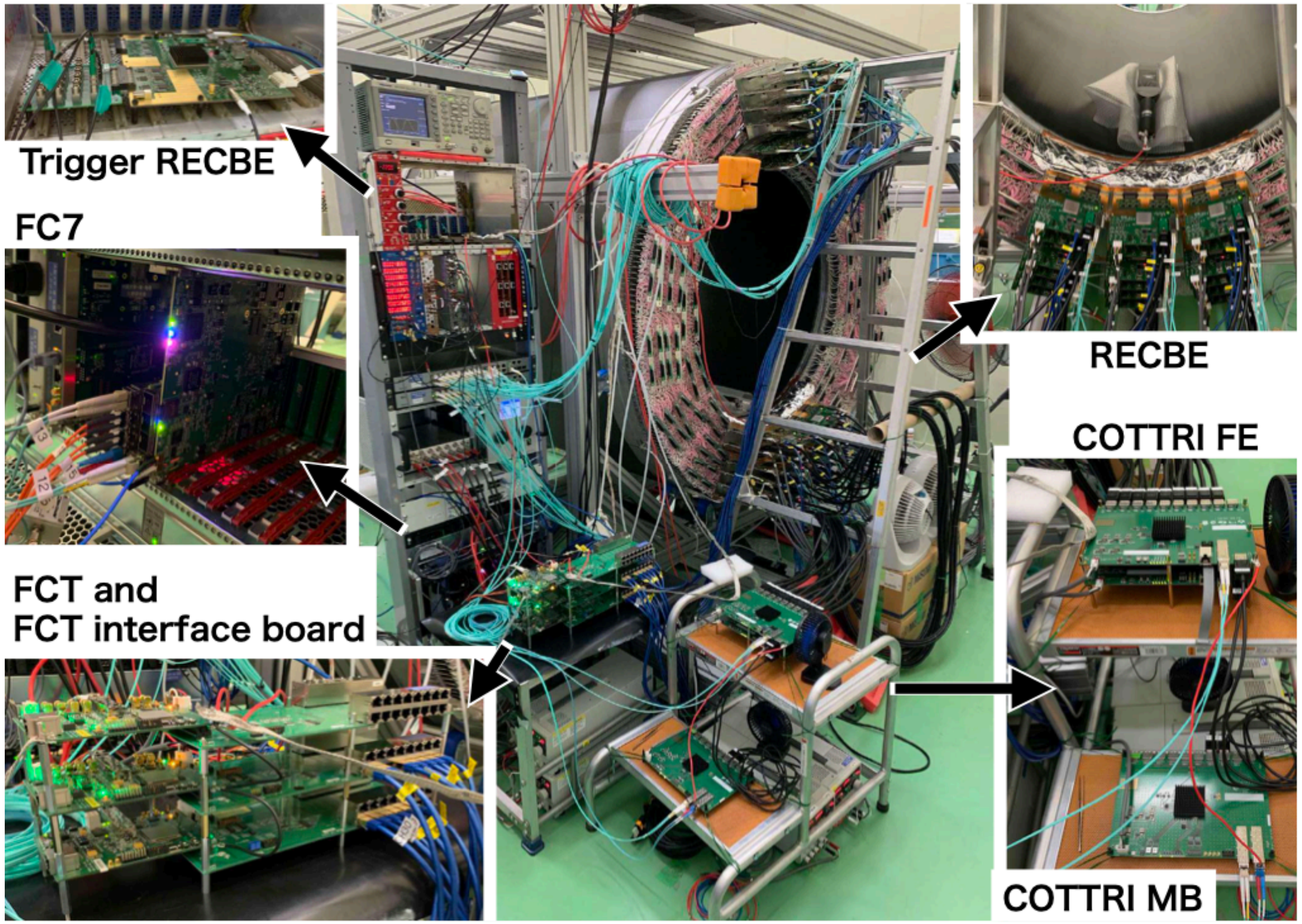
TRIGGER FULL CHAIN TEST

- Electronics full-chain test with a partial CDC
  - Incl. LUTs inside the FPGAs in COTTRI CTH FE for GBDTs
  - The cosmic-ray test was performed in 2019
  - Measured latency **3.2 μsec** obtained

	Latency [μs]	Description
RECBE - COTTRI System - FC7 - RECBE	1.9 - 2.0	100 ns fluctuation by the data transfer rate of 10 MHz
Drift time distribution	0.4	Data evaluation every 100 ns
Trigger receiving time in RECBE	0.8	32bit trig. data with 40 MHz







SUMMARY

- GBDT based online hit classification was proposed to extremely suppress the non-trajectory fake trigger
- Achieved a **96%** signal efficiency with less than **13 kHz** fake trigger rate from the original rate of ~90 kHz based on the simulation study
- A COTTRI system has been designed and full chain test was performed in success with the GBDT's LUTs already implemented
- Obtained **3.2 μsec** latency much shorter than the requirement of 7.5 μsec

See details in Y. Nakazawa's PhD thesis

SUMMARY

- GBDT based online hit classification was proposed to extremely suppress the non-trajectory fake trigger
- Achieved a **96%** signal efficiency with less than **13 kHz** fake trigger rate from the original rate of ~90 kHz based on the simulation study
- A COTTRI system has been designed and full chain test was performed in success with the GBDT's LUTs already implemented
- Obtained **3.2 μsec** latency much shorter than the requirement of 7.5 μsec
- **We want more!**
  - Further BG suppression → Wider timing window (**=larger signal acceptance**), **New bi-product trigger**, Sustainable data management, etc.

See details in Y. Nakazawa's PhD thesis

NEURAL NETWORK BASED EVENT CLASSIFICATION (1)

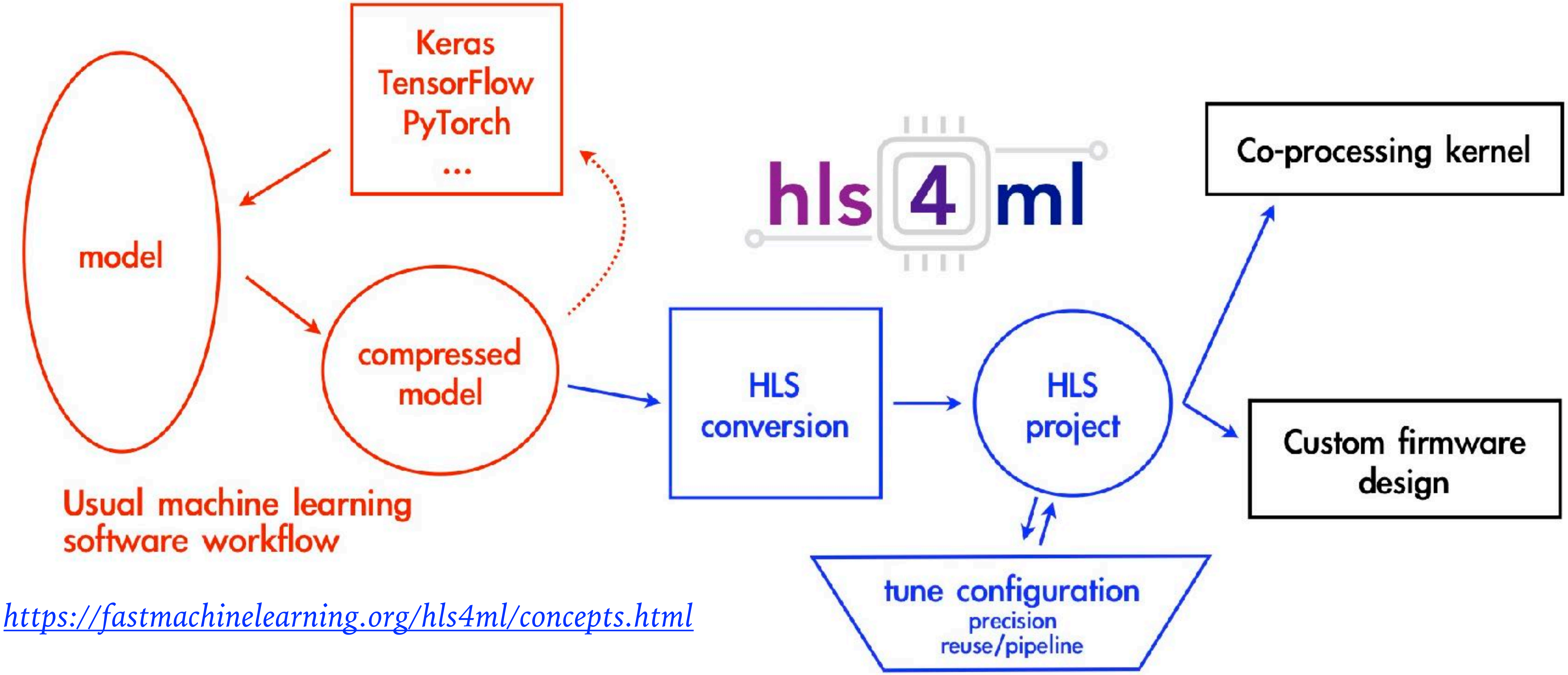
- NNs can be alternative (or additive) to the cut-based event classification after the GBDT hit classifier
  - **Pros**
    - Excellent pattern recognition capability especially with the deep neural networks
    - Various softwares available for the quick model evaluations
    - Much faster than the arithmetic calculations in general
  - **Cons**
    - Difficult model conversion from networks to the real firmware
    - Heavy resource usage (DSP/LUT/BRAM) for the calculation
    - Calibrations(?) uncertainty estimation(?)

NEURAL NETWORK BASED EVENT CLASSIFICATION (1)

- NNs can be alternative (or additive) to the cut-based event classification after the GBDT hit classifier
  - Pros
    - Excellent pattern recognition capability especially with the deep neural networks
    - Various softwares available for the quick model evaluations
    - Much faster than the arithmetic calculations in general
  - Cons
    - ~~Difficult model conversion from networks to the real firmware~~ → *New tools available (hls4ml)*
    - ~~Heavy resource usage (DSP/LUT/BRAM) for the calculation~~ → *Sparse networks with model quantisations*
    - Calibrations(?) uncertainty estimation(?) → *Not to be covered today*

MODEL CONSTRUCTION (1)

➤ General workflow of the NN development for FPGA using ***hls4ml***



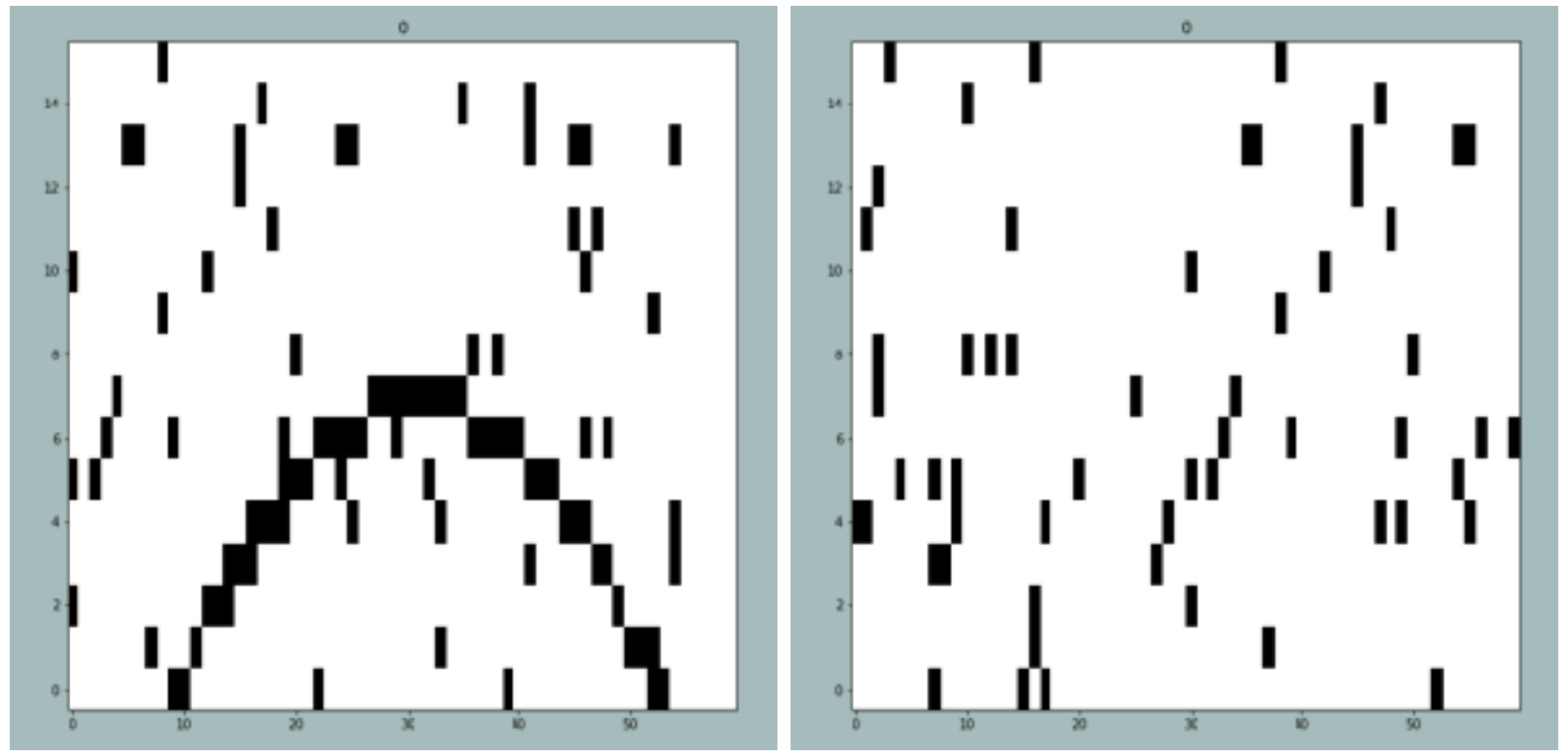
https://fastmachinelearning.org/hls4ml/concepts.html

MODEL CONSTRUCTION (2)

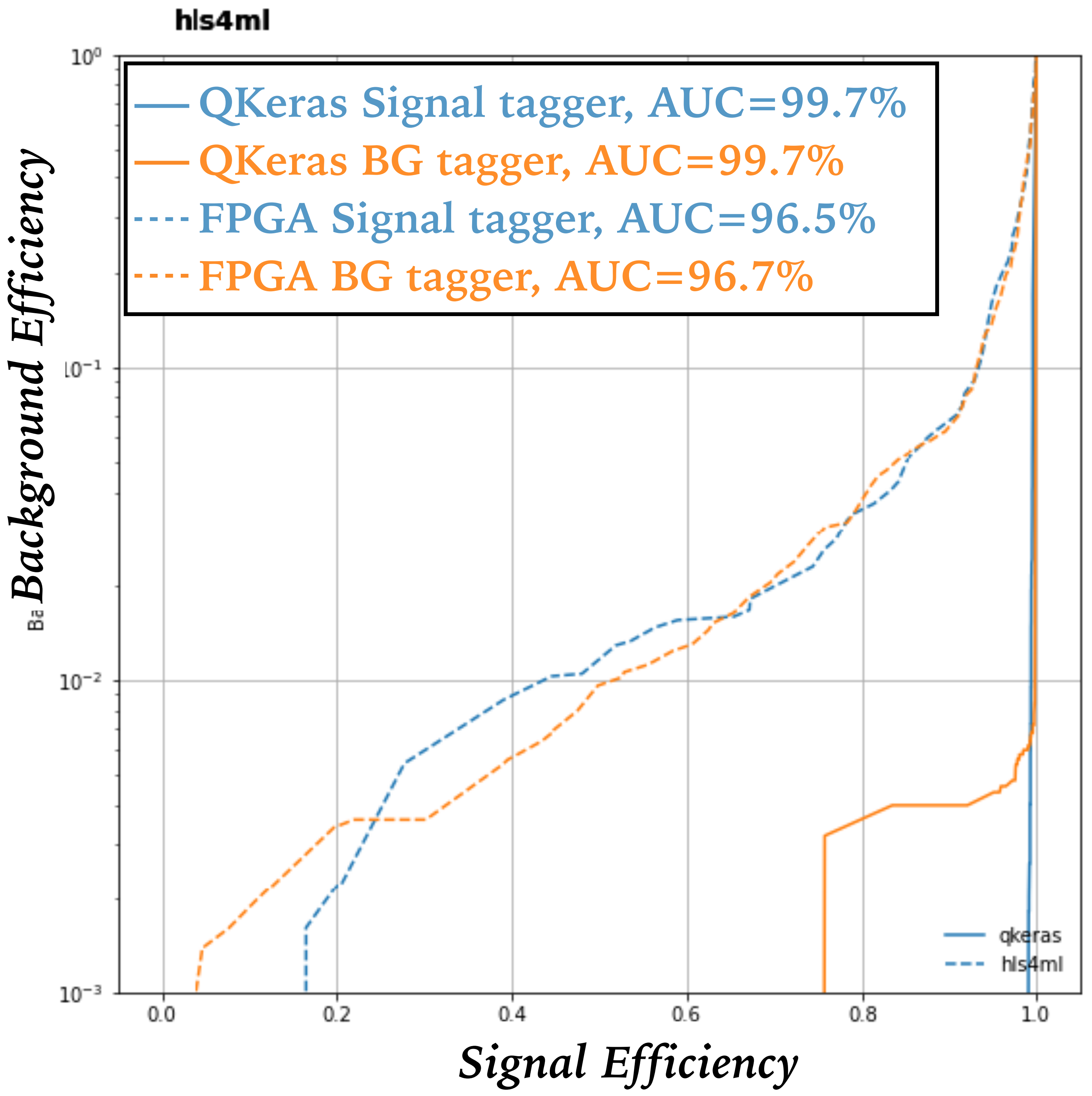
- What do we (users) do (in general)?
  1. Data preparations and formatting
  2. Model selections
  3. Parameters' tuning (# of layers, sparseness, resolutions etc.)
     ➡ Grid scanning, built-in/customised tuners, etc.
  4. Performance evaluation
     ➡ Accuracy, latency, stability etc
  5. Resource check
     ➡ Select your FPGA chip and see whether resource is available

MODEL CONSTRUCTION (3)

- As a first test, we made sets of toy MC for signal/background events for NN training/test
  - 5% noise events randomly distributed with/without the arch (signal-like) pattern
- Quantised and sparse Multi layer perceptron (QMLP) was tentatively chosen
  - Few hyper-parameters tuned roughly by utilising a Keras built-in Bayesian optimiser



Signal *BG*



Resource usage @Kintex-7 xc7k355T-FFG901 (%)

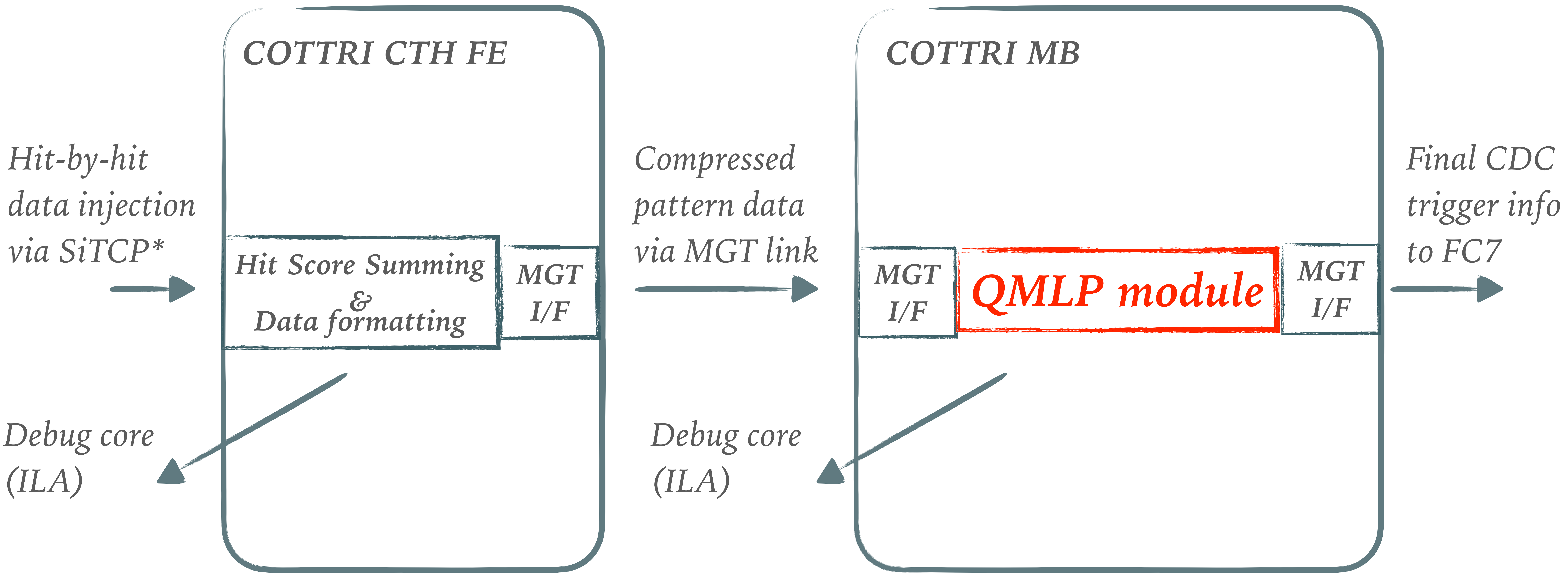
BRAM	DSP	FF	LUT
0	0	5	32

Latency estimated to be 260 clock cycles

= 130ns @200MHz

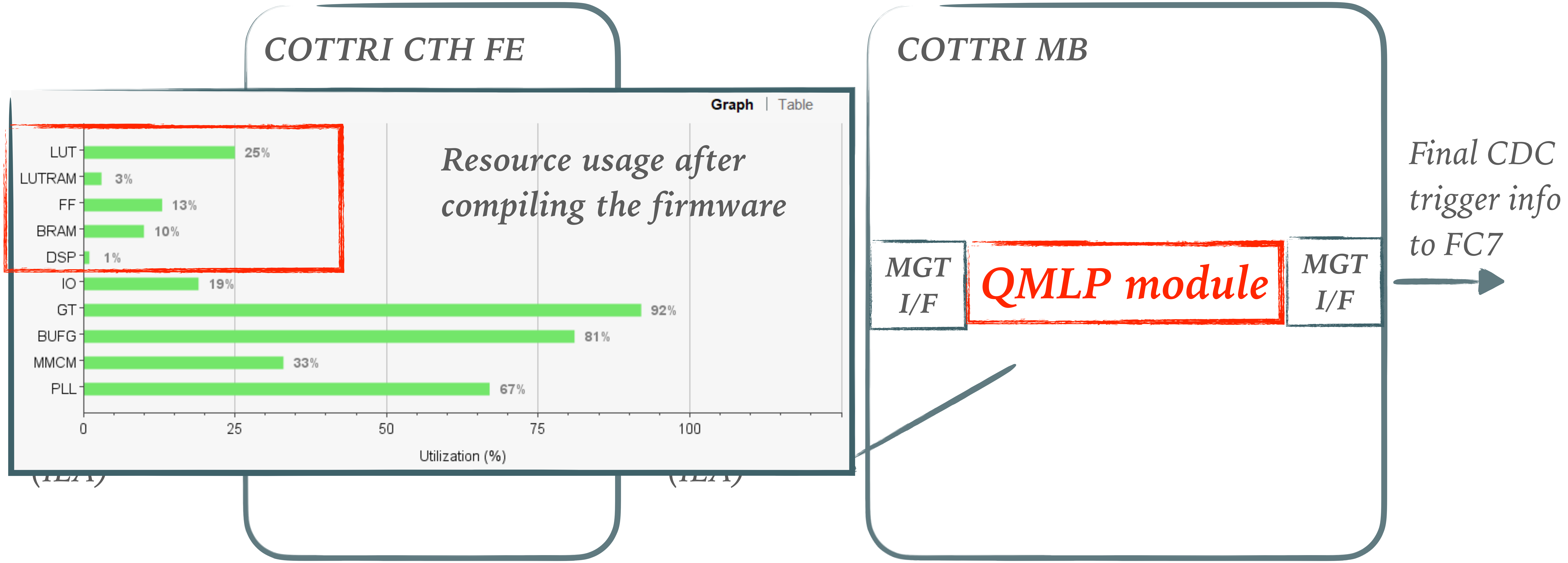
FIRMWARE DEVELOPMENT (1)

- Structure of the "test" firmware



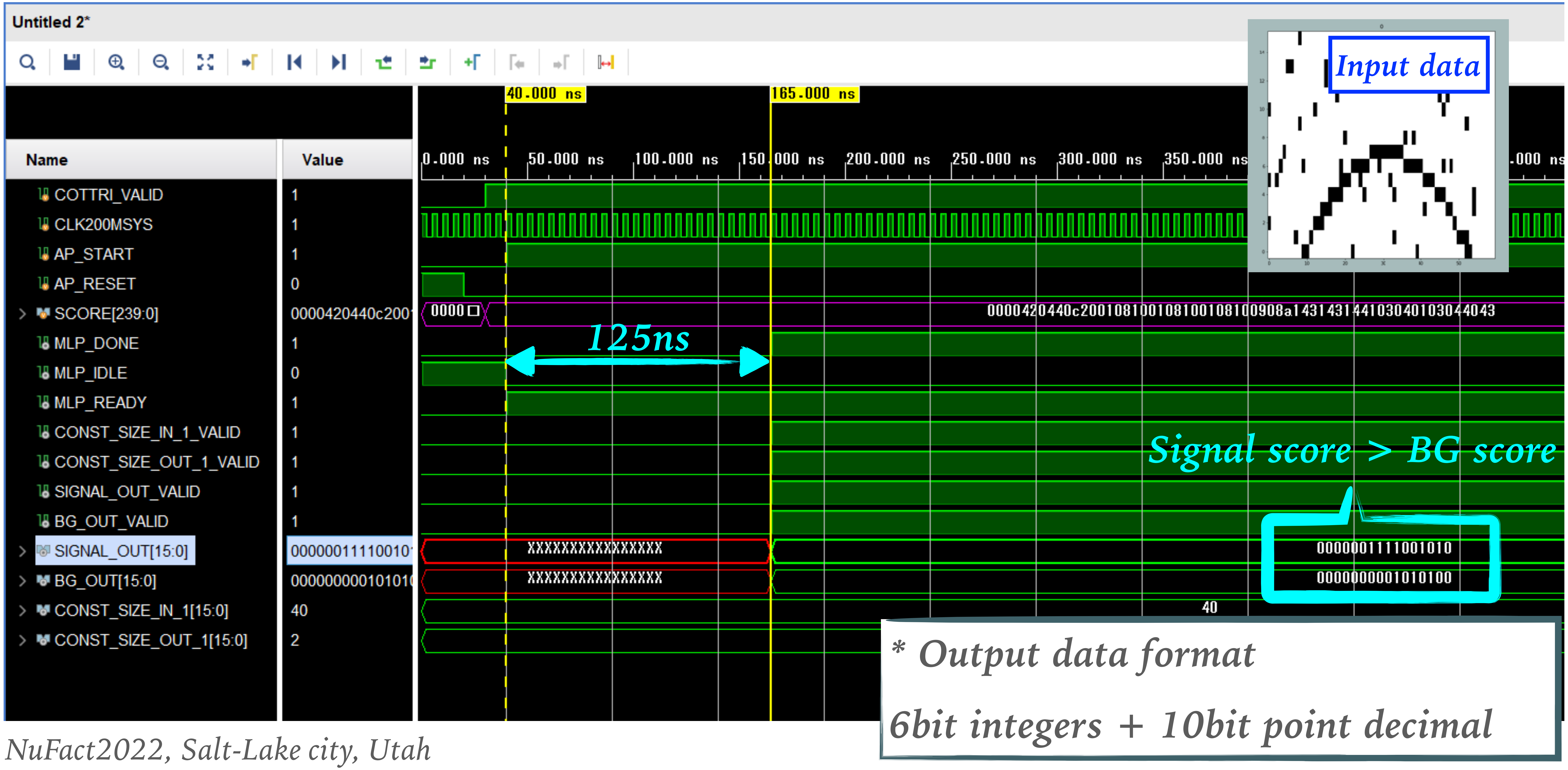
FIRMWARE DEVELOPMENT (1)

- Structure of the “test” firmware



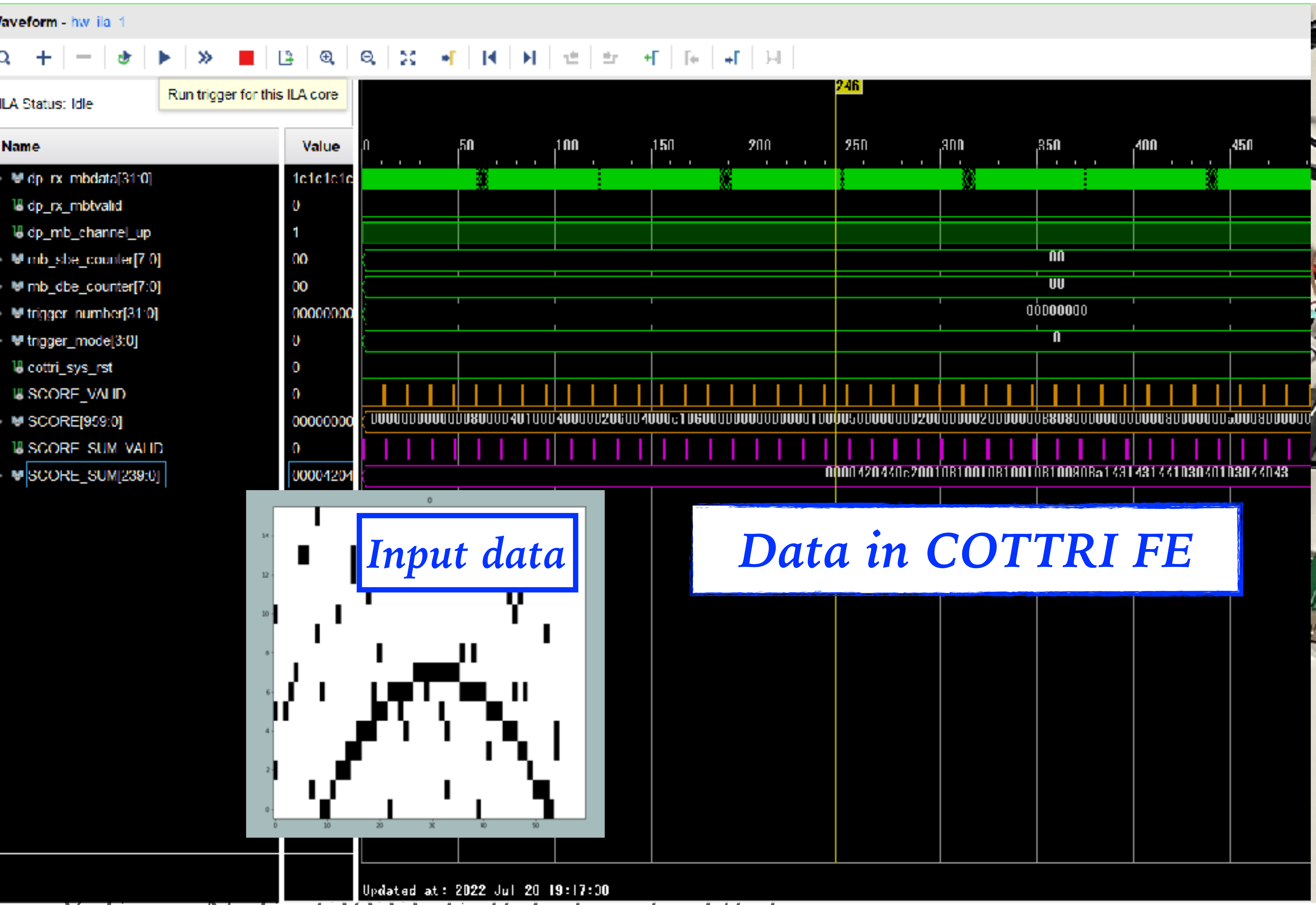
FIRMWARE DEVELOPMENT (2)

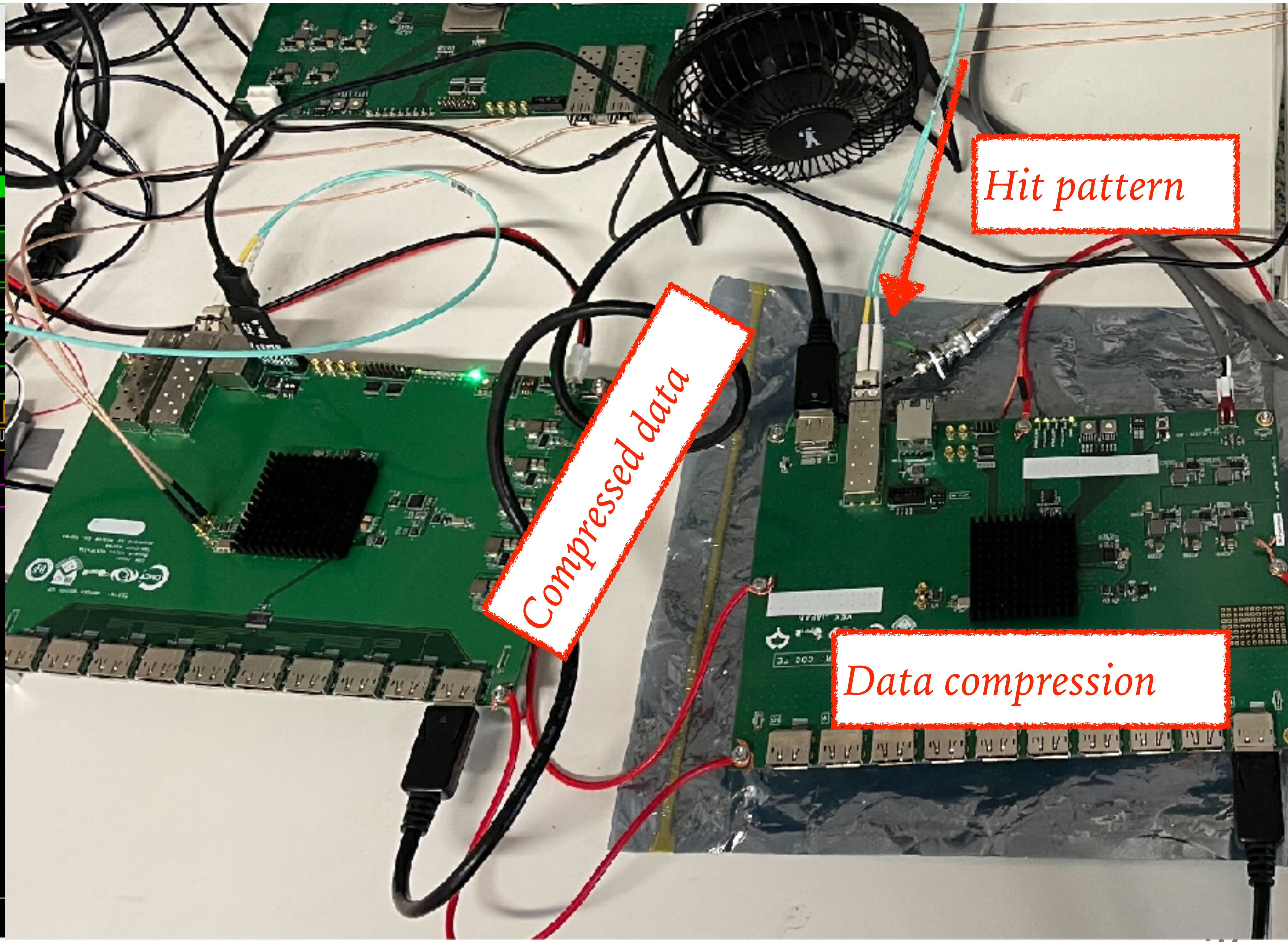
- Firmware simulation with Vivado



HARDWARE TEST (1)

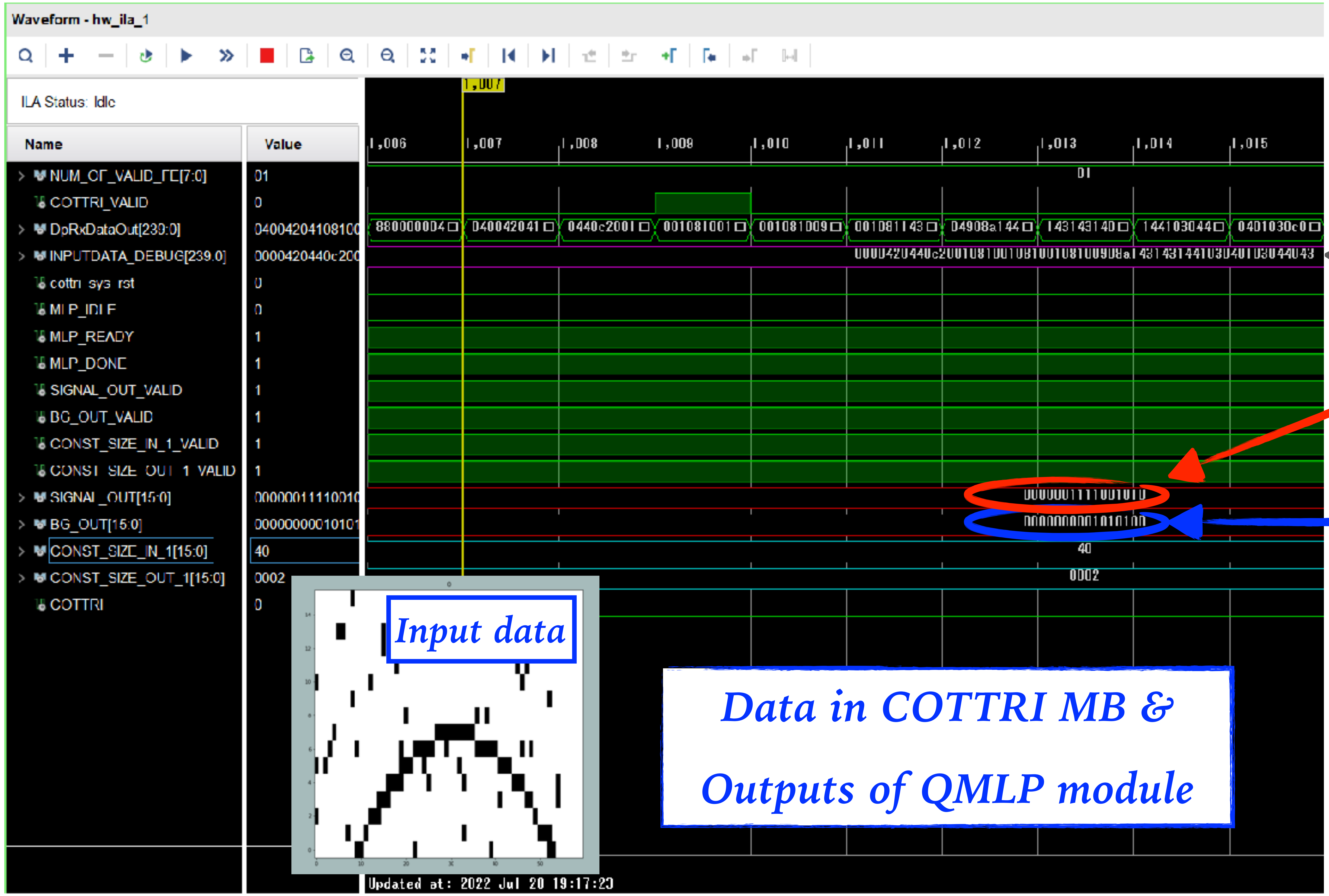
- Actual NN firmware module (QMLP) was implemented into the COTTRI MB
  - Write MC signal/BG data pattern into FE via UDP protocol & send them to MB via **2.4 Gbps** MGT link
  - NN classification performed inside the FPGA & outputs were checked by using **Vivado ILA debug core**





HARDWARE TEST (2)

➤ NN firmware implementation results (just obtained in the last week!)



Received data from COTTRI FE

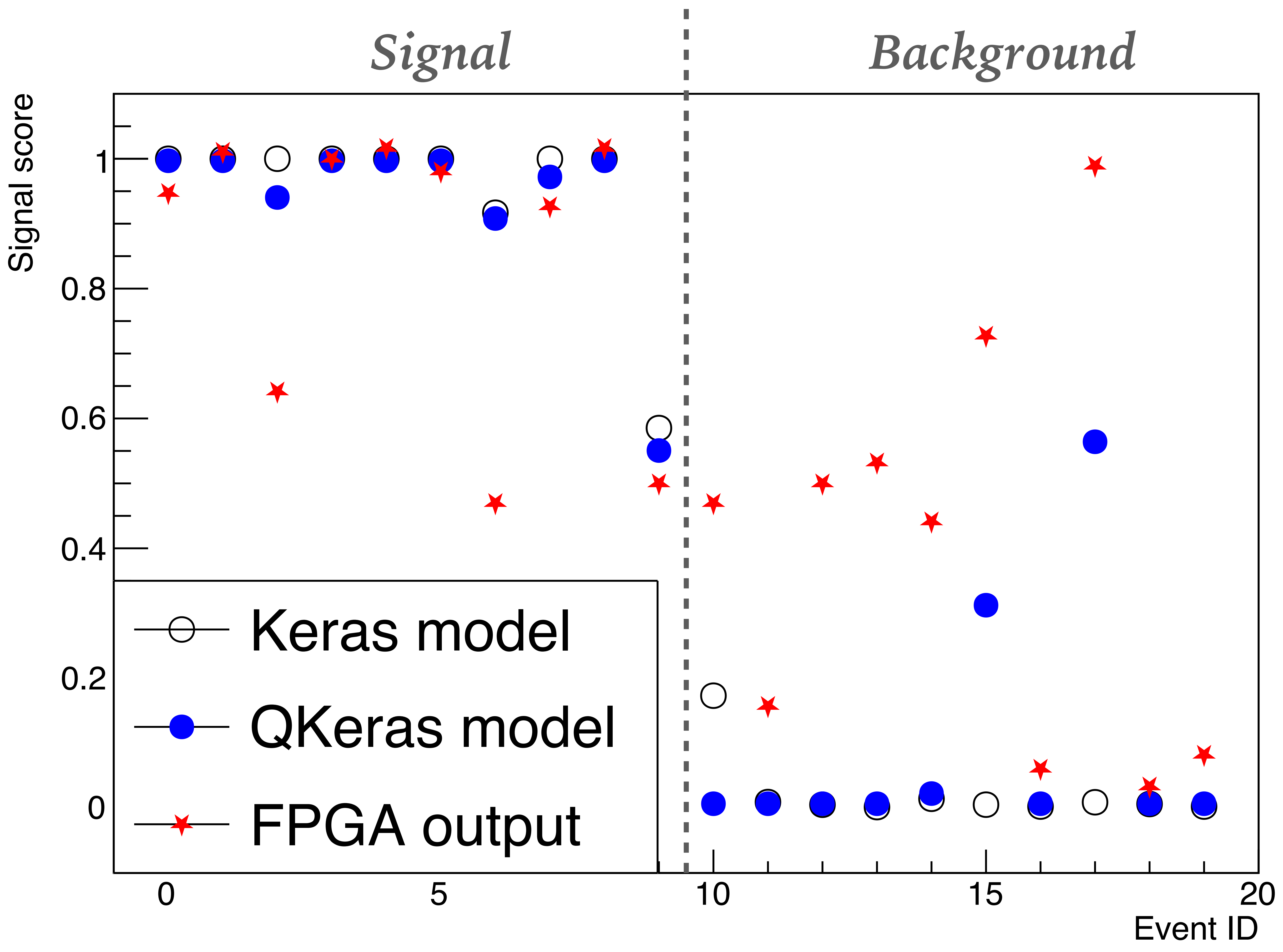
Output of Signal Classifier ~0.95

Output of BG Classifier ~0.08

Both output values are consistent with both simulation and offline outputs

RESULTS

- Comparisons for10 signal/BG events
  - Online classifier shows similar but worse performance compared to the offline MLP models
  - More events to be checked
- This is a very preliminary test in order to establish the workflow of NN implemented FPGA
  - More resources available
  - Further optimisations available



SUMMARY AND PROSPECTS

- A fast and highly efficient trigger is essential in the COMET Phase-I experiment
  - **Better trigger, more physics**
- Online machine learning algorithms inside FPGAs are being developed
  - GBDT based hit classification was developed and the simulation study showed **96%** signal efficiency + **13 kHz** trigger rate with a very short net-latency, **3.2 μsec**
- Additional NN based event classification was proposed and the development has begun
  - Potential increasing of the signal sensitivity by factor of two
  - Sparse QMLP model can be realised with very low FPGA resources
  - We established the workflow and **the NN-based firmware was designed, generated and tested with a real FPGA board**

Thank you!

BACK UP

- QMLP model structure
  - Very sparse model was chosen for the first trial