Accelerating Machine Learning algorithms in FPGAs for the trigger system of a SiPM-based upgraded camera of the CTA Large-Sized Telescopes
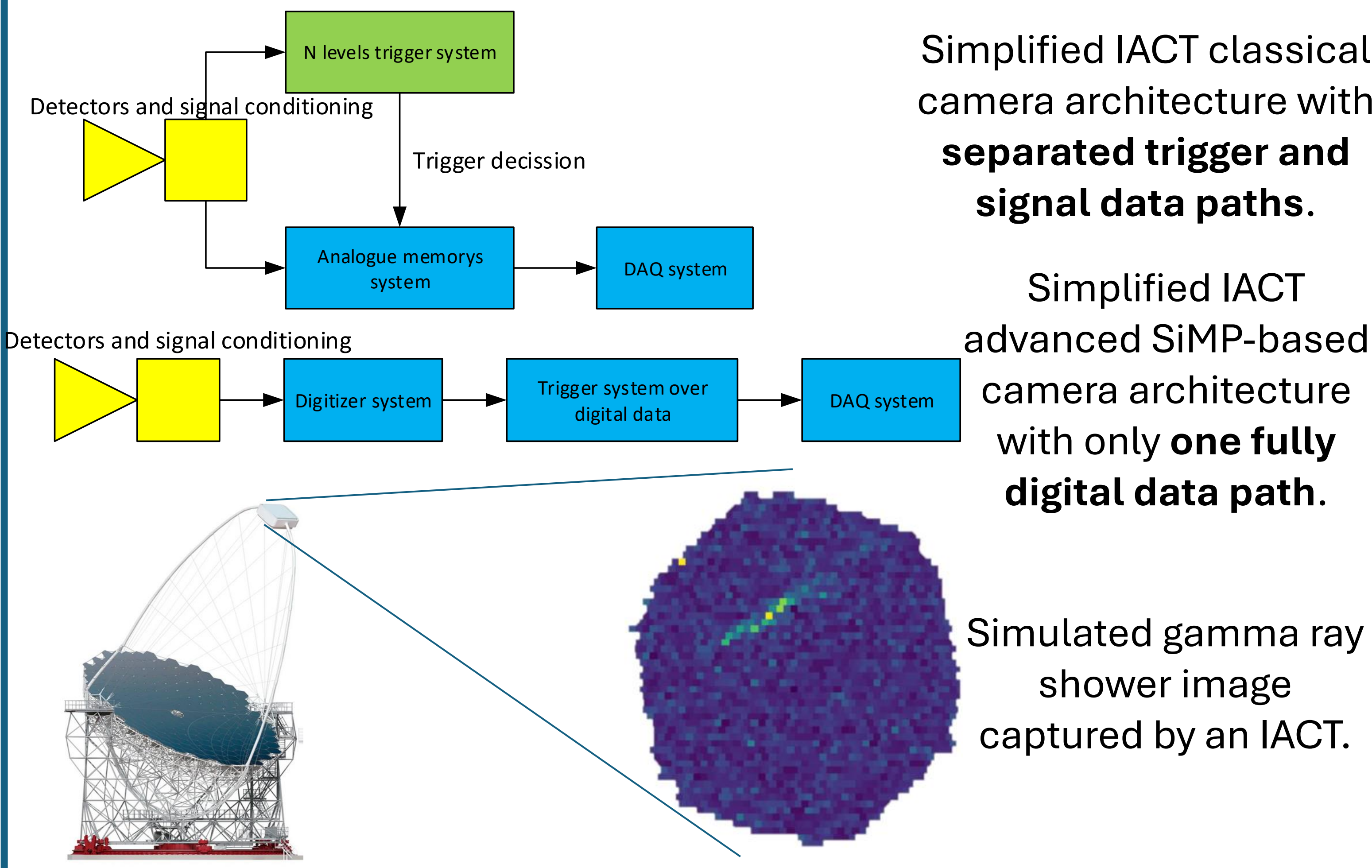
Alejandro Pérez Aguilera (1), L.A. Tejedor (1), J.A. Barrio (1), T. Miener (2), D. Martín (1)

(1) Grupo de Altas Energías (GAE), Instituto de Física de Partículas y del Cosmos, and EMFTEL Department, Universidad Complutense de Madrid (IPARCOS-UCM), E-28040 Madrid, Spain
(2) University of Geneva - Département de physique nucléaire et corpusculaire, 24 Quai Ernest Ansernet, 1211 Genève 4, Switzerland

Abstract

Current Imaging Atmospheric Cherenkov Telescopes use combined analog and digital electronics for their trigger systems, implementing simple but fast algorithms. Such trigger techniques are used due to high data rates and strict timing requirements. In recent years, in the context of a future upgraded camera for the Large-Sized Telescopes (LSTs) of the Cherenkov Telescope Array (CTA) based on Silicon PhotoMultipliers, a new fully digital trigger system incorporating Machine Learning (ML) algorithms is being developed. The main aim is to implement those algorithms in FPGAs to increase the sensitivity and efficiency of the real-time decision making while being able to fulfill timing constraints. The project is full of challenges, such as complex printed circuit board design, complex FPGA logic design, and translating high level ML models to FPGA synthesizable code. We are currently developing a test bench as a proof of concept and to evaluate the FPGA performance of the algorithms.

IACT camera architectures

Simplified IACT classical camera architecture with **separated trigger and signal data paths**.

Simplified IACT advanced SiMP-based camera architecture with only **one fully digital data path**.

Simulated gamma ray shower image captured by an IACT.

Requirements Overview

- The direct output data of the camera are **hexagonal shaped events** that need to be processed in real time to tag/eliminate as much only-noise (Night Sky Background) events as possible.
- The event rate after the L1 trigger will be around **hundreds of kHz**.
- The event data can be divided into **Regions of Interest (RoI)**.
- Processing time should be in the **few µs** range ➔ FPGAs

Software tools

- IACT images can be processed offline with **CNNs** defined and trained with **CTLearn** [2] software package.
- **hls4ml** software package is used to create FPGA firmware of ML algorithms, CNNs in this case.
- With Xilinx Vivado software the IPs generated by hls4ml [3] are integrated into a firmware project an evaluated with a test bench.

Simulations show that shower/NSB separation is possible with this small models in the trigger system [1].

Contact: alejpe15@ucm.es

Latency and utilization estimates

- The VHDL test bench is fed by RoIs composed **of 5 samples of 30x30 pixel images**.
- The part chosen is **Kintex UltraScale** tier.
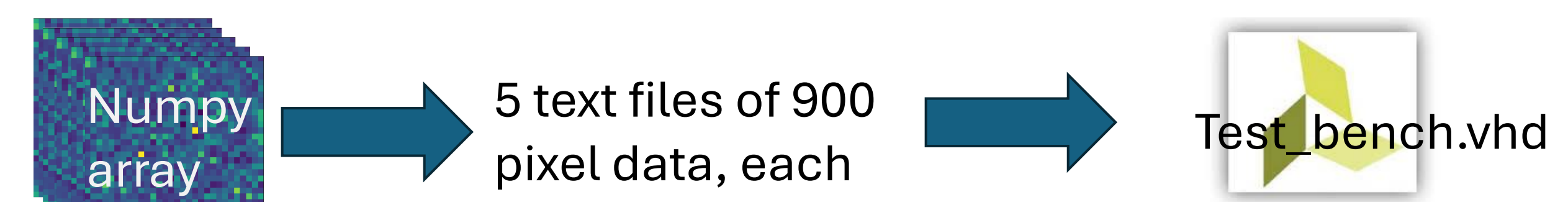
- Only the option of Reuse Factor 1 is compliant with the requirements with a 5 ns clock.
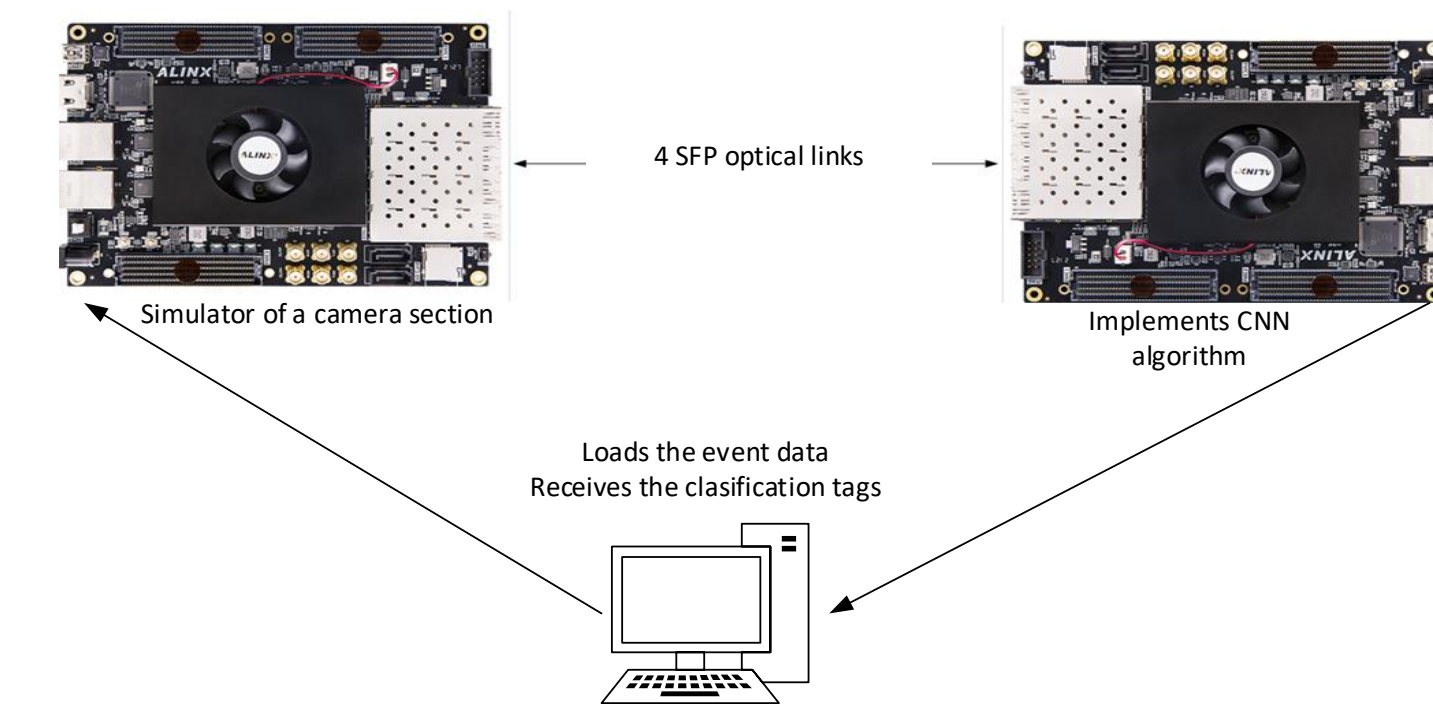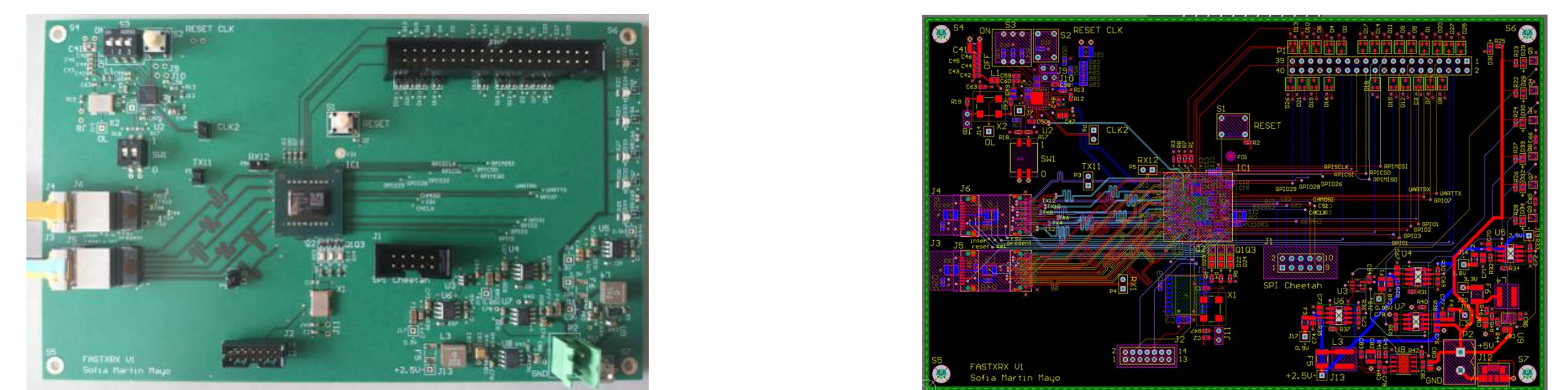- This model size fits into the desired tier of FPGA.

R. Factor	Latency (us)	DSP
1	5.2	122
8	12.9	66
16	15.3	52
32	15	29
64	20.4	17
128	33	9
256	41	6

Hardware for testbench and prototypes

- 2 development boards with Kintex UltraScale tier FPGAs (Alinx XCKU040 with 4 SFPs). One to simulate a camera section, the other to implement the CNN algorithm.

- PCB manufacturing to test the physical interfaces: Firefly tx/rx, Artix UltraScale+, MicroWave PCB substrate

Discussion and near future activities

- **Several RoIs** need to be processed **in parallel** to cover all the area of a camera event.
- Further optimizations of the CNN models, such as **quantization aware training**, are yet to be explored.
- Density-Based Scan models also to be explored.
- Works to check the **tagging performance** is ongoing.
- Recently joined DRD7.5 WP to share expertise.
- Short-term: test-bench/algorithms characterized by 2026.
- Mid-term: full prototype produced by 2028.

Acknowledgements

References

[1] I. Bezshyiko, C. Abellán Beteta, et al."Deep Learning-Based Data Processing in Large-Sized Telescopes of the Cherenkov Telescope Array Observatory: FPGA Implementation," in EuCAIFCon, Amsterdam, 2024.
[2] CTLearn (2024) https://github.com/ctlearn-project/ctlearn, DOI: 10.5281/zenodo.11475531
[3] T. Aarrestad, et al., "Fast convolutional neural networks on FPGAs with hls4ml" *Mach. Learn. Sci. Tech.*, vol. 2, 045015, Jul 2021