CTAO

Machine-Learning-based CTAO Telescope data processing

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

Agencia Estatal de Investigación

Comunidad de Madrid

Astrofísica y Física de Altas Energías

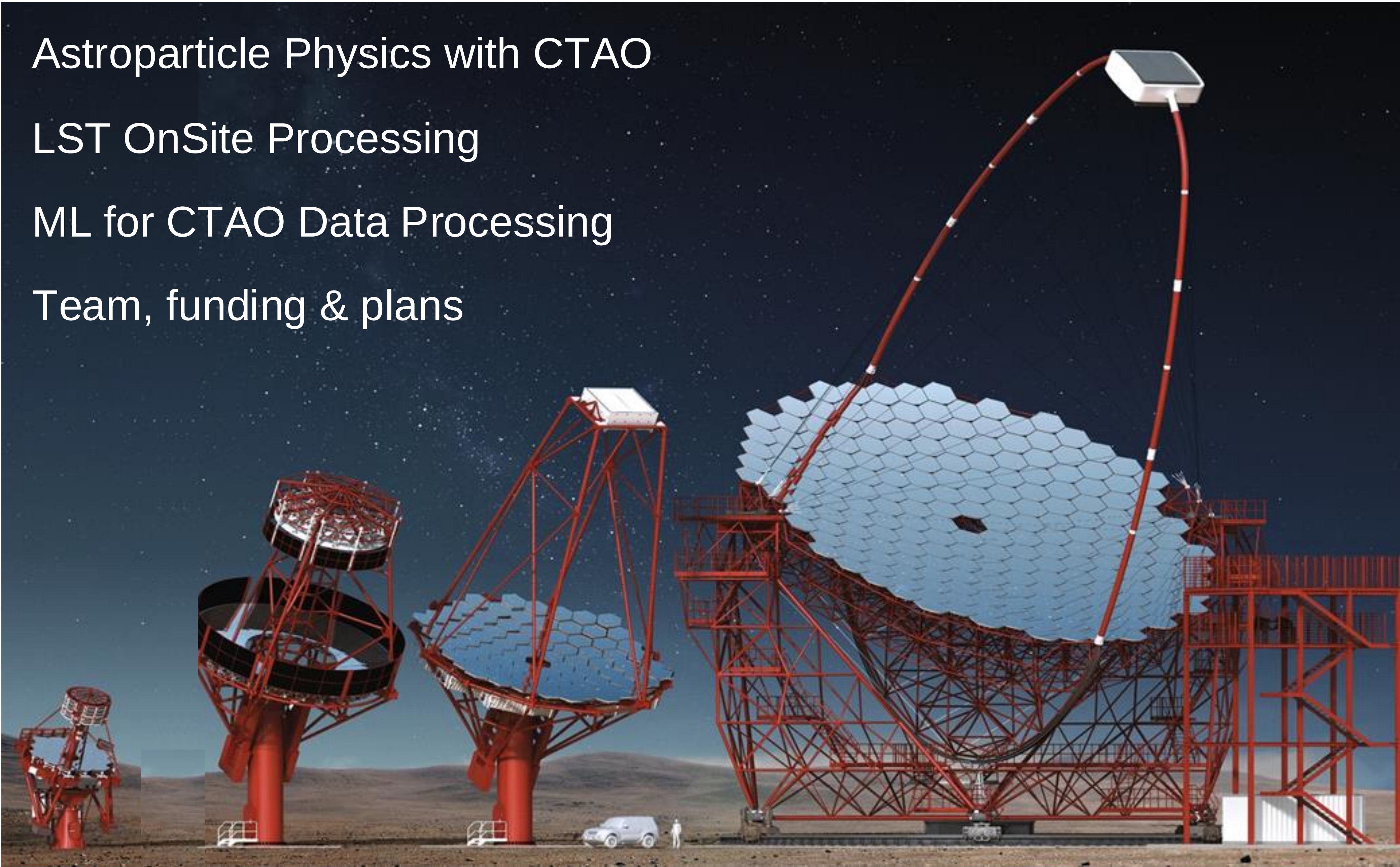
COMCHA 2nd Workshop – October 2024

A. Pérez-Aguilera, M. Lainez, A. Cerviño, D. Martín,
L.A. Tejedor, D. Nieto, J.L. Contreras, J.A. Barrio

GAE IPARCOS Universidad Complutense Madrid

High Energy Physic Group – Institute of Particles and Cosmos Physics – University Complutense of Madrid

Contents



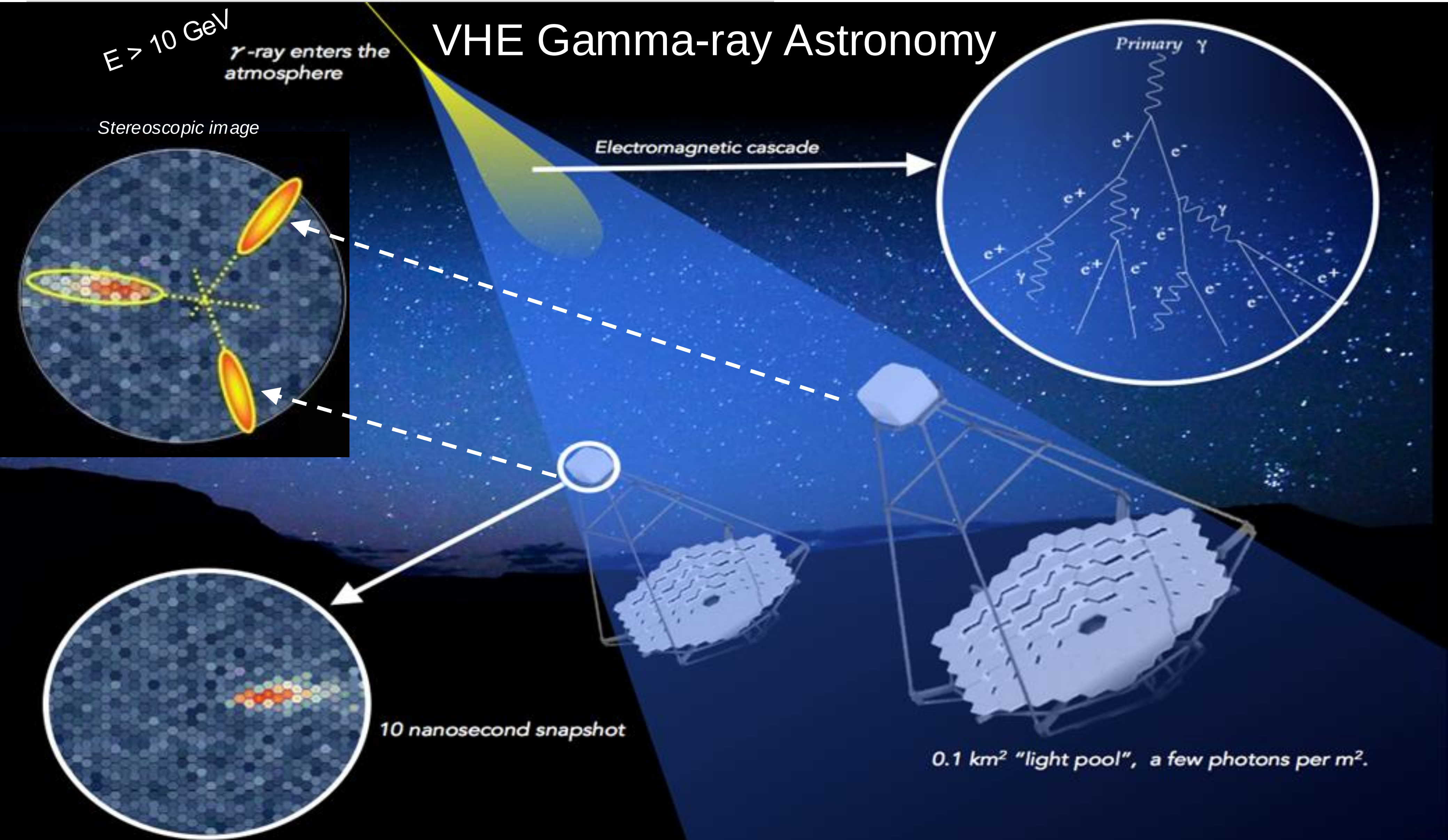
Astroparticle Physics with CTAO

LST OnSite Processing

ML for CTAO Data Processing

Team, funding & plans

Astroparticle Physics



Imaging Atmospheric Cherenkov Telescopes (IACTs) → *Homogeneous EM Calorimeter*

CTA Observatory

Cherenkov Telescope Array Observatory

Sensitivity improvement x10
Energy range extension x10
Angular resolution improvement

→

Two observatories:
La Palma (*Canaries*) / Chile
~100 telescopes

CTA Observatory

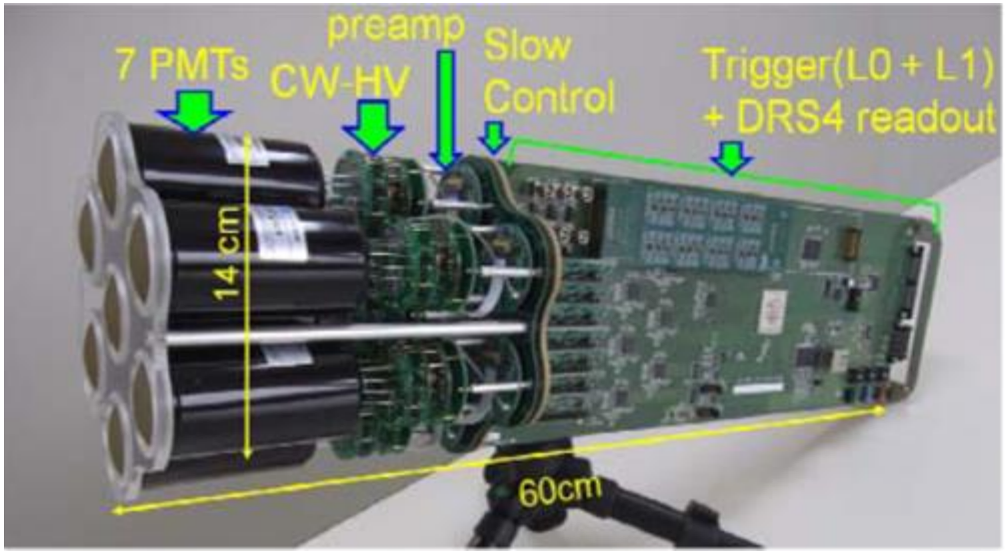
Cherenkov Telescope Array Observatory

Sensitivity improvement x10
Energy range extension x10
Angular resolution improvement

→

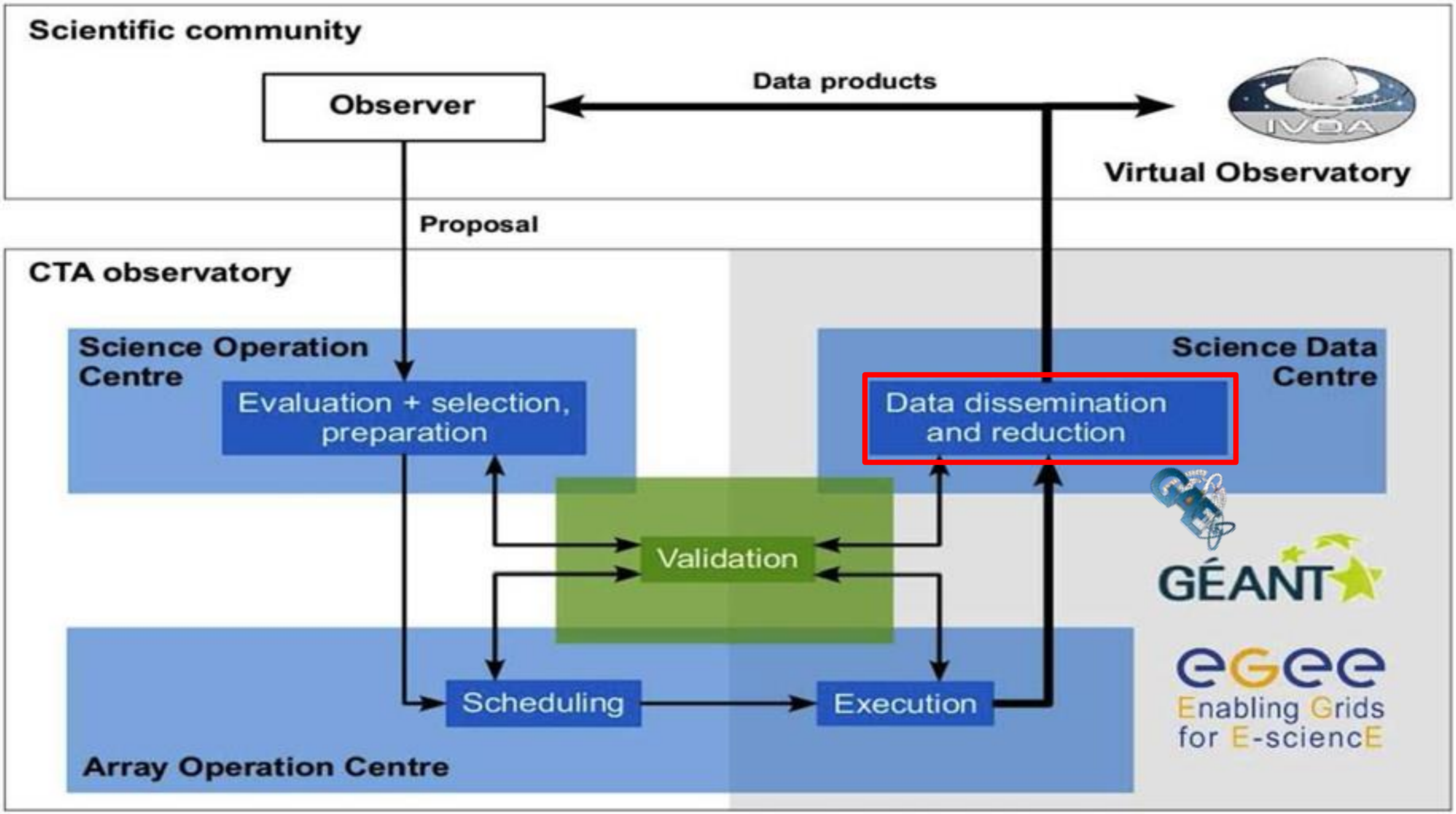
Two observatories:
La Palma (*Canaries*) / Chile
~100 telescopes

Cherenkov Telescope Array Observatory



~2000-pixel PMT-based camera

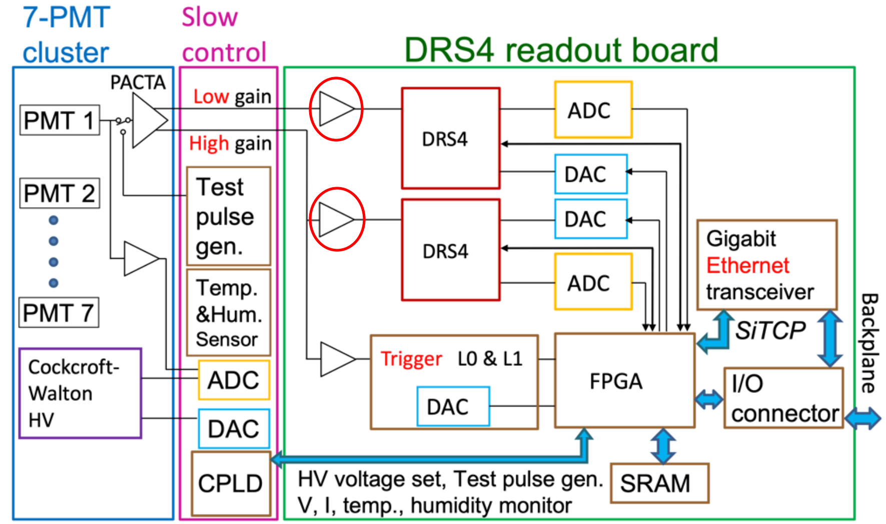
Cherenkov Telescope Array Observatory



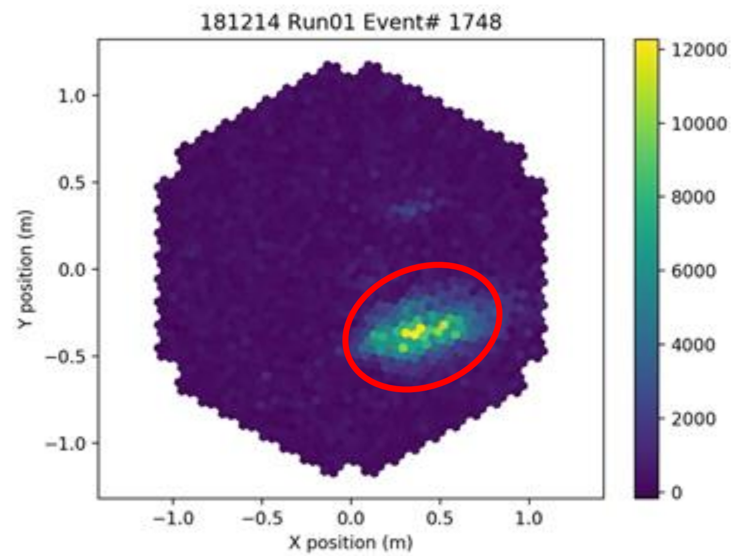
Data Processing @ CPUs

LST OnSite Processing for Data Volume Reduction

- OnSite Processing
  - LST1 alone produces ~10-20 TB/night raw data; 3 more LSTs to enter commissioning in ~2025
  - Data is processed on-site in a *temporary* data center at the telescope base, with ~1800 cores and 5.3 PB HD
  - Onsite pipeline process raw data every morning in a highly parallel way
  - Building on MAGIC OnSite experience → until recently saving all raw data



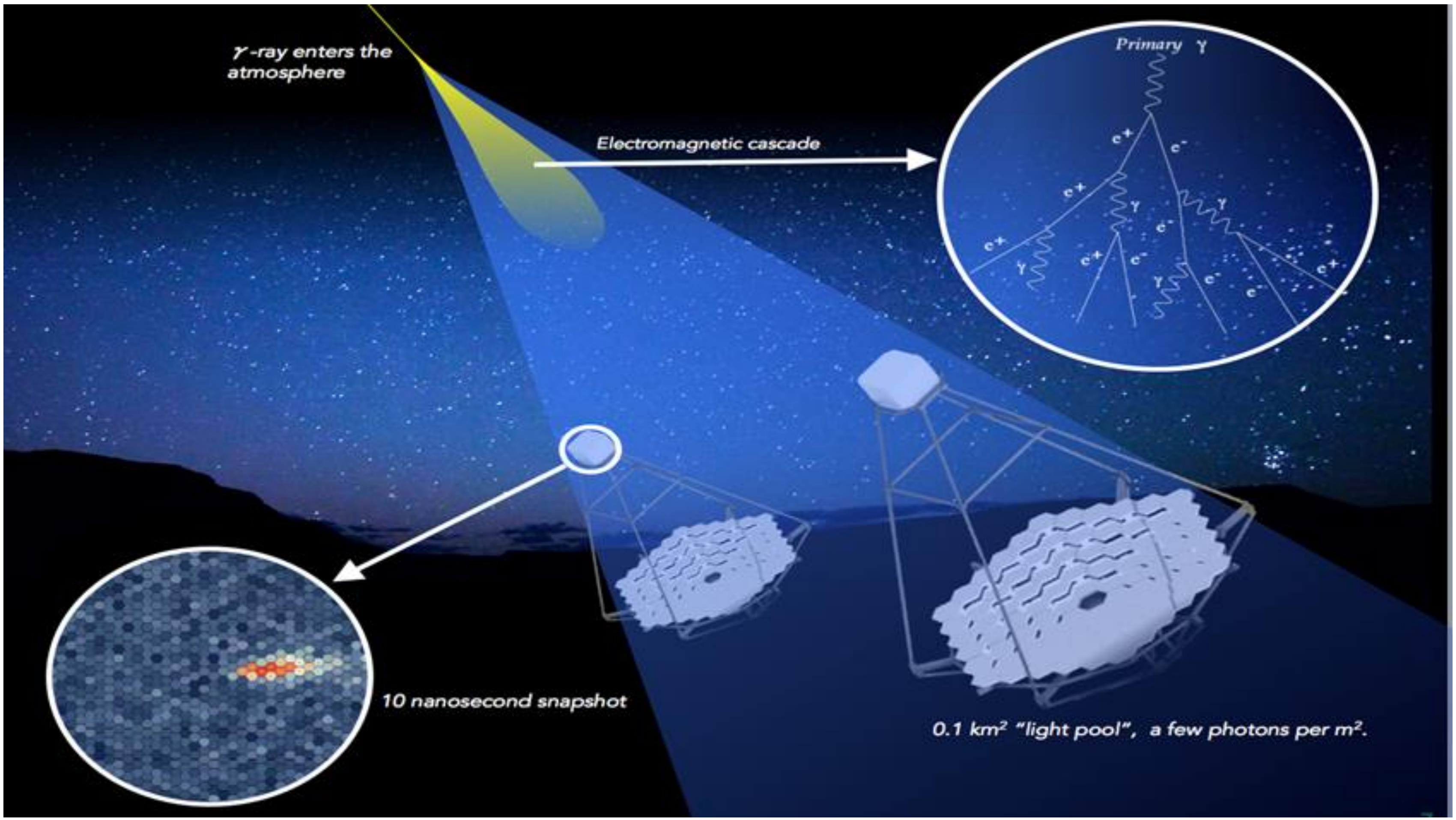
- Data Volume Reduction @ OnSite pipeline
  - 1st step: select only one of the 2 PMT-amplifier gains
  - 2nd step: Region of Interest selection, along with UAH
  - LST as test-bed for CTAO DVR



ML@GPUs for CTAO Reco

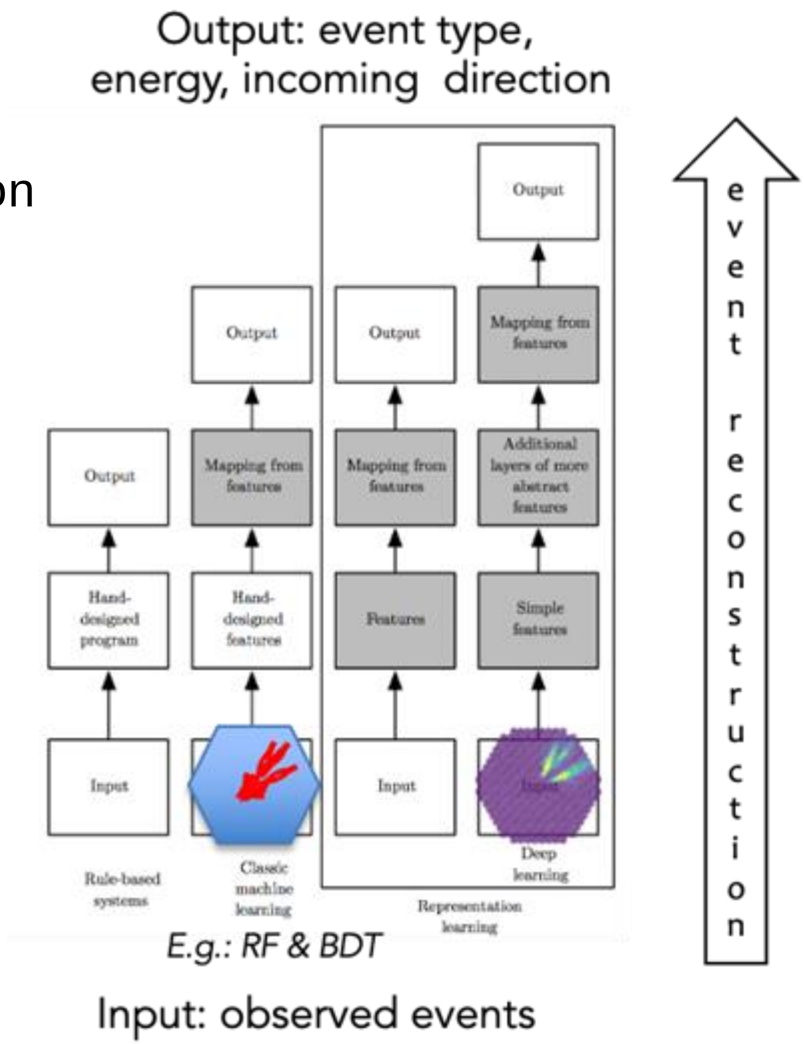
CTLearn

- High-level Python package for using deep learning at IACT:
  - gamma/hadron/Night-Sky-Background separation



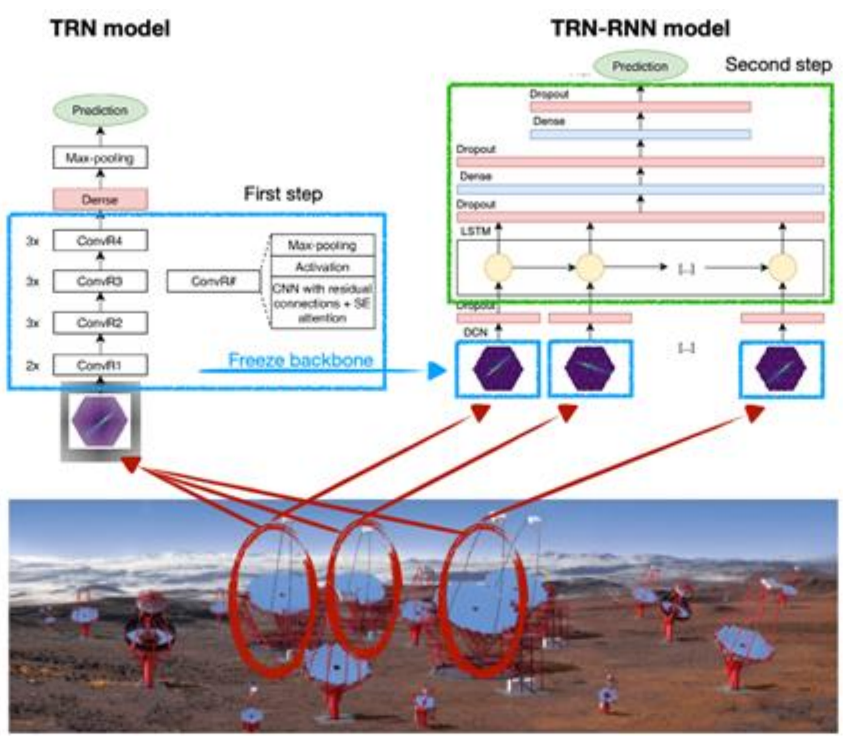
CTLearn

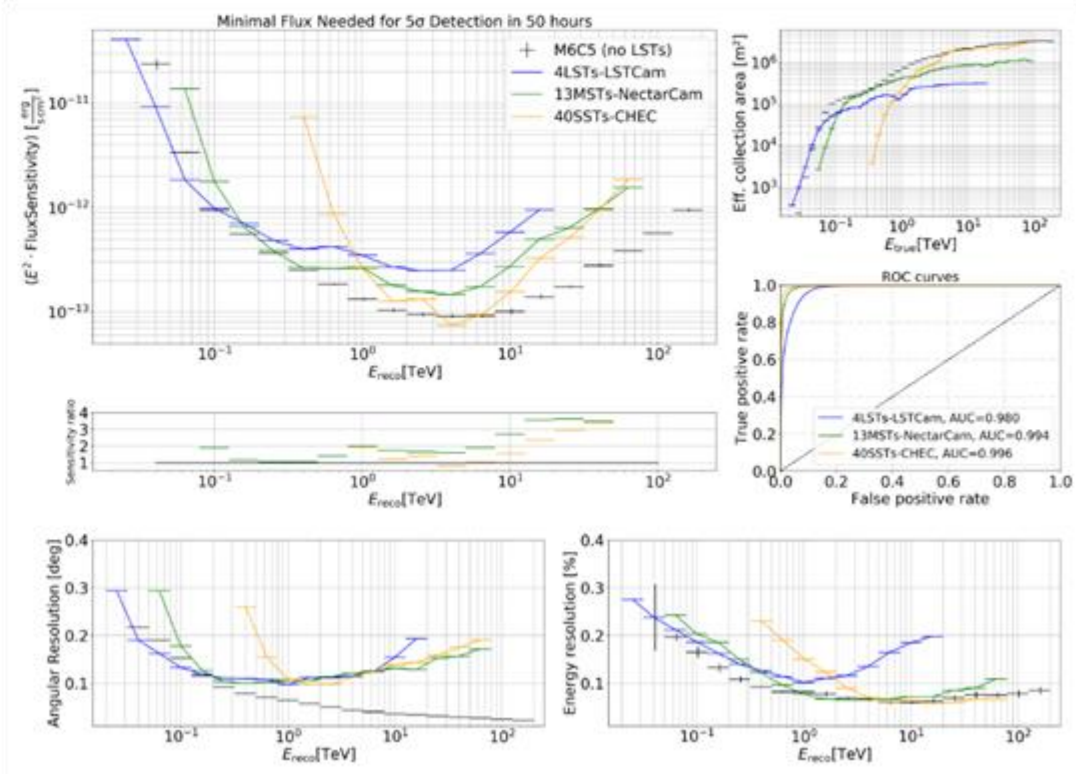
- High-level Python package for using deep learning at IACT:
  - gamma/hadron/Night-Sky-Background separation
  - Event Reconstruction
  - ➢ Based of pixel information
- Founded at GAE, led by GAE and UniGe
- Open source:
  https://github.com/ctlearn-project/ctlearn
  https://zenodo.org/records/11475531



ML@GPUs for CTAO Reco

CTLearn *on simulation*

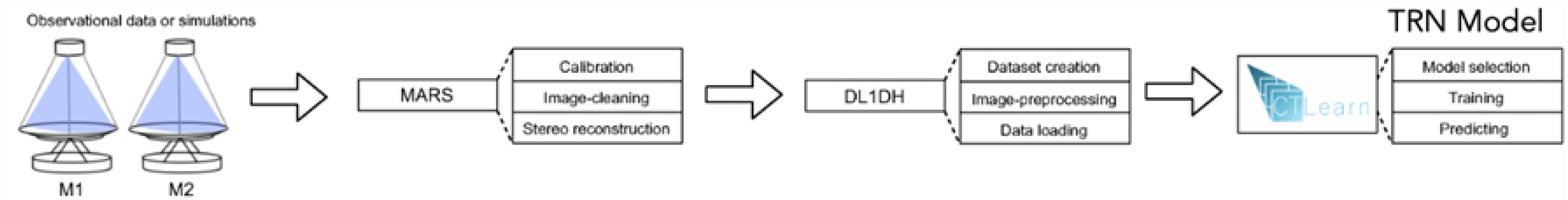


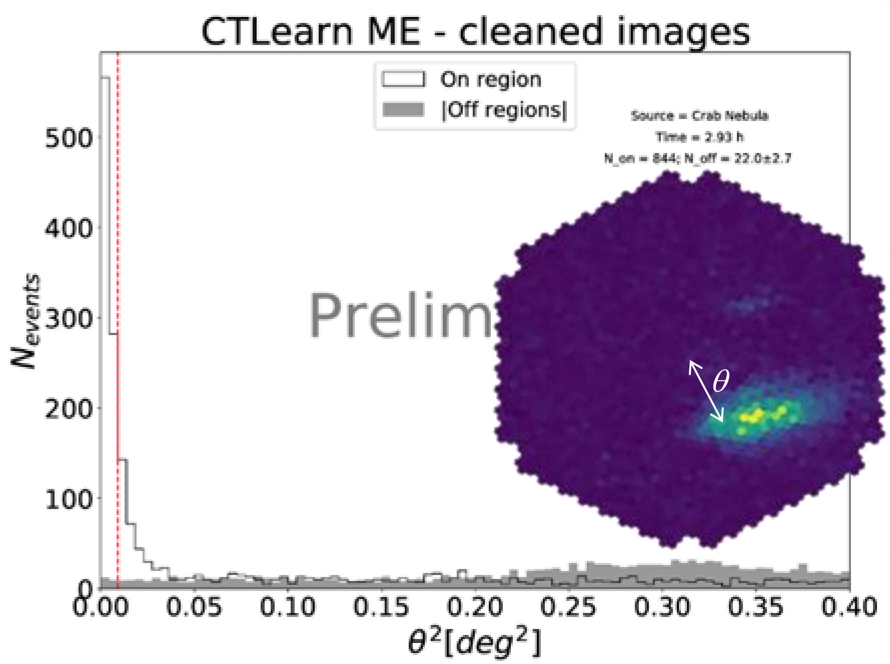


T. Miener et al., PoS(ICRC2021) 730

Application to CTAO simulated data demonstrated !!

CTLearn *on real data*





Analysis	γ rate [/min]	bkg rate [/min]	Sen. [% Crab]	Sig. (Li&Ma)
MARS – ME	4.54 ± 0.16	0.119 ± 0.015	0.70 ± 0.05	43.0σ
CTLearn – ME (raw)	3.45 ± 0.14	0.133 ± 0.018	0.97 ± 0.08	36.5σ
CTLearn – ME (cleaned)	4.68 ± 0.17	0.125 ± 0.015	0.69 ± 0.05	43.6σ
MARS – LE	16.49 ± 0.35	3.861 ± 0.086	1.09 ± 0.03	61.1σ
CTLearn – LE (raw)	11.70 ± 0.32	3.832 ± 0.114	1.53 ± 0.05	47.5σ
CTLearn – LE (cleaned)	16.24 ± 0.35	3.872 ± 0.086	1.11 ± 0.03	60.4σ

Analysis	N_{on}	N_{off}	N_{ex}
MARS – ME	819	21.0 ± 2.6	798.0 ± 28.7
CTLearn – ME (raw)	629	23.3 ± 3.1	605.7 ± 25.3
CTLearn – ME (cleaned)	844	22.0 ± 2.7	822.0 ± 29.2
MARS – LE	3579	679.0 ± 15.0	2900.0 ± 61.7
CTLearn – LE (raw)	2730	673.7 ± 20.0	2056.3 ± 56.0
CTLearn – LE (cleaned)	3536	680.7 ± 15.1	2855.3 ± 61.3

Summary of all performed analyses of the same Crab Nebula sample

T. Miener et al. 2021 (ADASS XXXI)

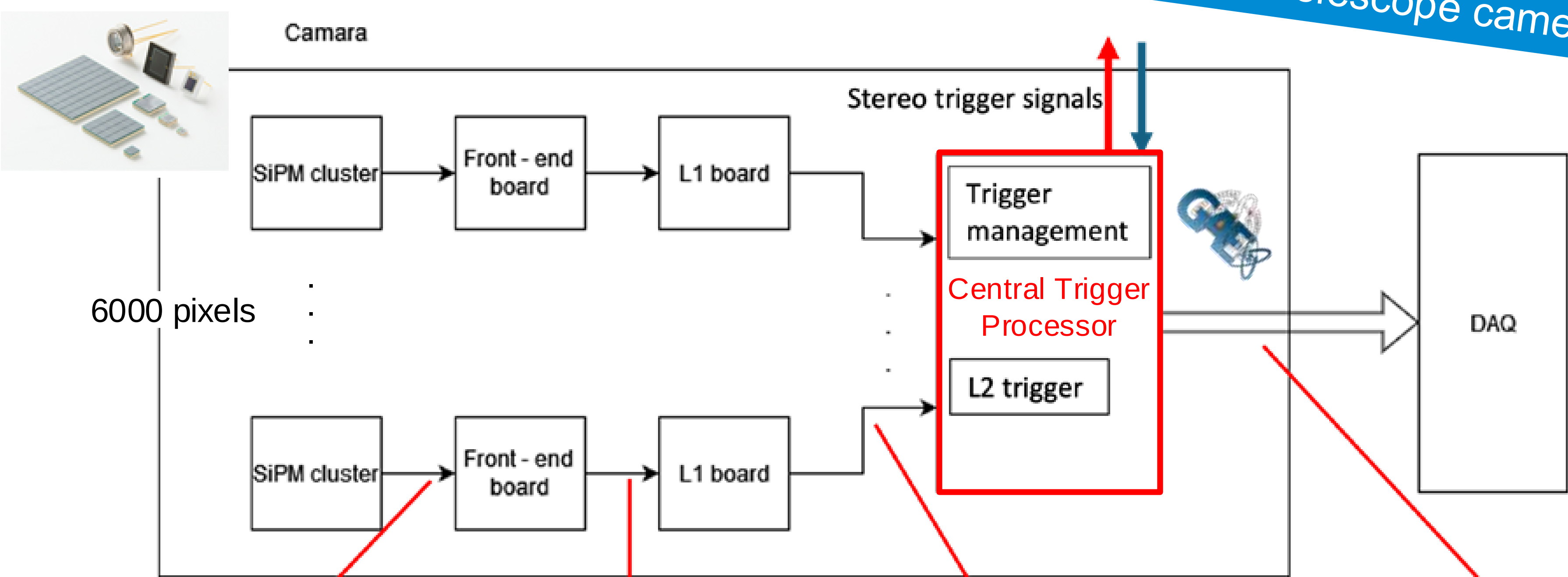
Application to MAGIC real data demonstrated !!

LST1 real data coming soon

ML@FPGAs for CTAO Trigger

Advanced LST SiPM Camera*

Candidate for mid-term upgrade of CTAO telescope cameras



Raw analog SiPM signal

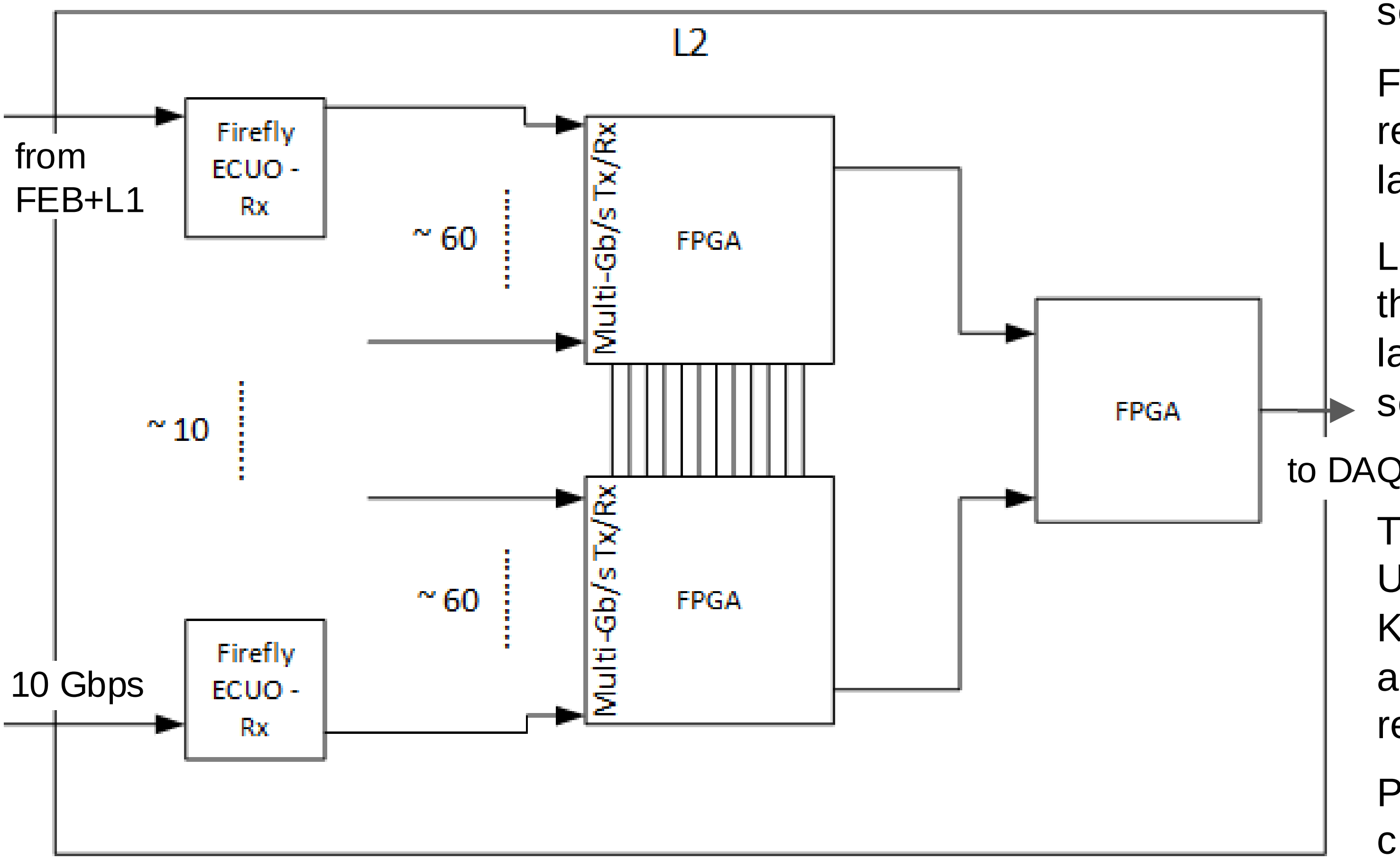
Digital SiPM signal
1 GHz sampling rate
~72 Tb/s

Full event after L1 trigger
300 kHz* event rate
10Gb/s optical links

Full event after L2 trigger
30 kHz* event rate

*M. Heller et al. PoS(ICRC2023)740

Central Trigger Processor: ML-based L2 trigger on FPGAs



L2 Trigger O(1 μs) latency

FPGAs in layers to allow scalability

First layer of FPGAs for data reception, formatting and first layers of CNN algorithm

Last FPGA will take care of the CNN last fully connected layers, Event Building and sending the data to DAQ.

Tentative FPGA models: Kintex UltraScale KU085, KU095 or KU115, with moderate cost and a large number of high-speed resources.

Prelim. number of high-speed channels, still under design

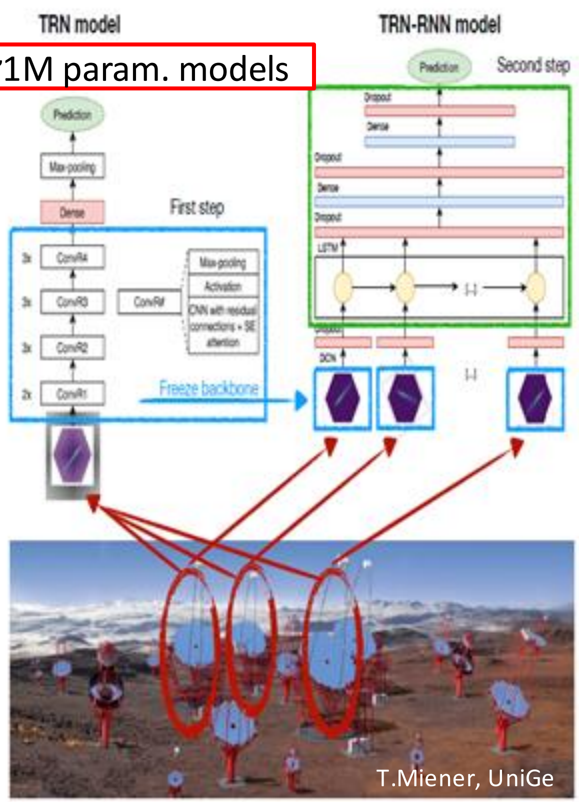
ML@FPGAs for CTAO Trigger

ML algorithms for CTAO camera trigger

- Train small CNNs & optimize large CNNs to fit in FPGAs
  → promising shower/NSB separation

- Translating CNN models to firmware code

~1M param. models



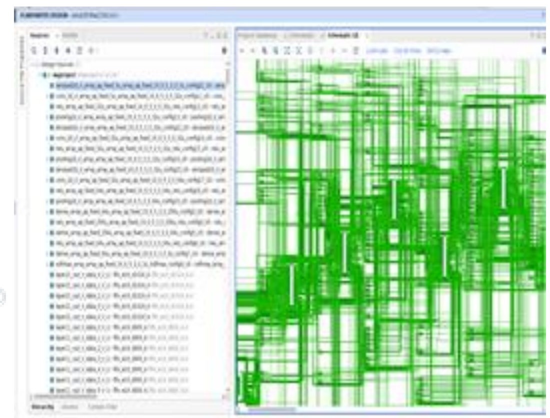
~2k param. models

Model: "CTLearn_model"

Layer (type)	Output Shape	Param #
waveforms (InputLayer)	[(None, 30, 30, 5)]	0
SingleCNN_block (Functional)	(None, 16)	1536
fc_particletype_1 (Dense)	(None, 32)	544
particletype (Dense)	(None, 3)	99
type (Softmax)	(None, 3)	0

Total params: 2179 (8.51 KB)
Trainable params: 2179 (8.51 KB)
Non-trainable params: 0 (0.00 Byte)

hls4ml



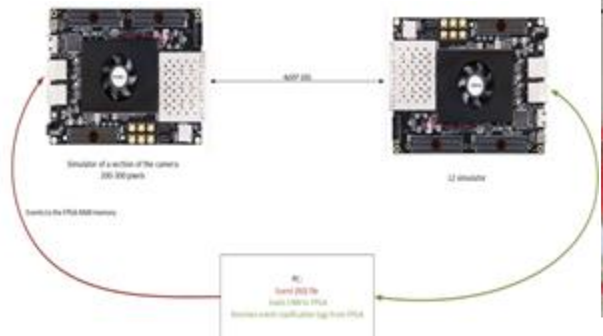
Verilog code

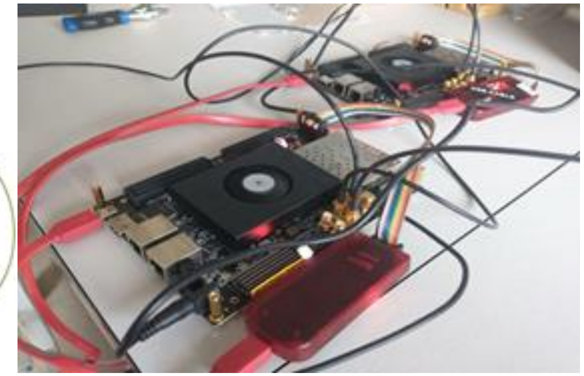
Vivado IP

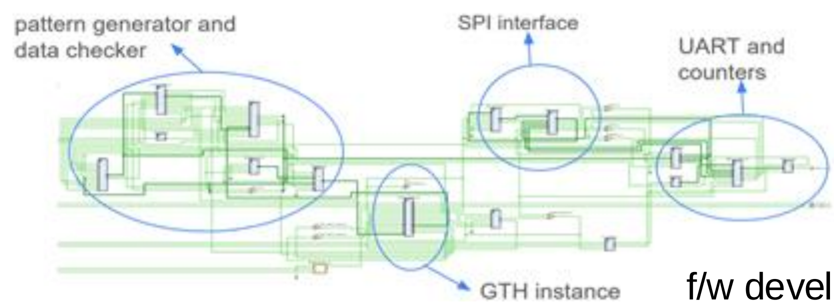
ML@FPGAs for CTAO Trigger

Prototype 1: CNN@FPGA trigger testbench

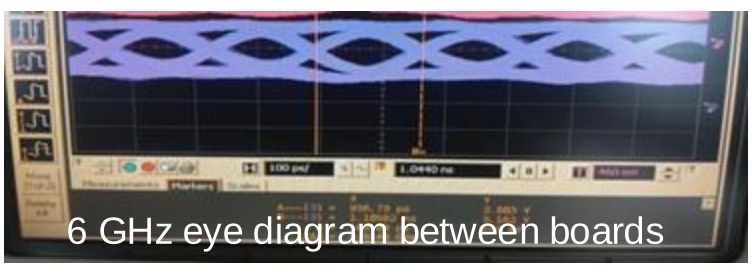
- Building the testbench
  - 2 development boards with Kintex UltraScale tier FPGAs (Alinx XCKU040 with 4 SFPs)
  - One board to simulate pixel data information (i.e. FEB+L1) and the other to implement & test firmware core and L2 algorithms
- Using the testbench
  - Develop f/w for multi-gigabit transceivers and other interfaces.
  - Test f/w in the development boards.



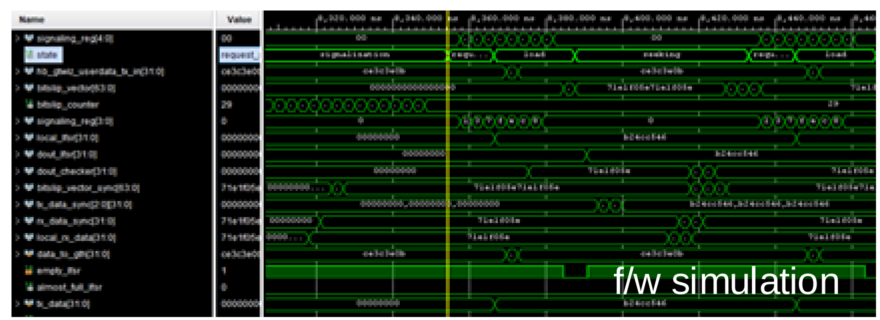




f/w development



6 GHz eye diagram between boards

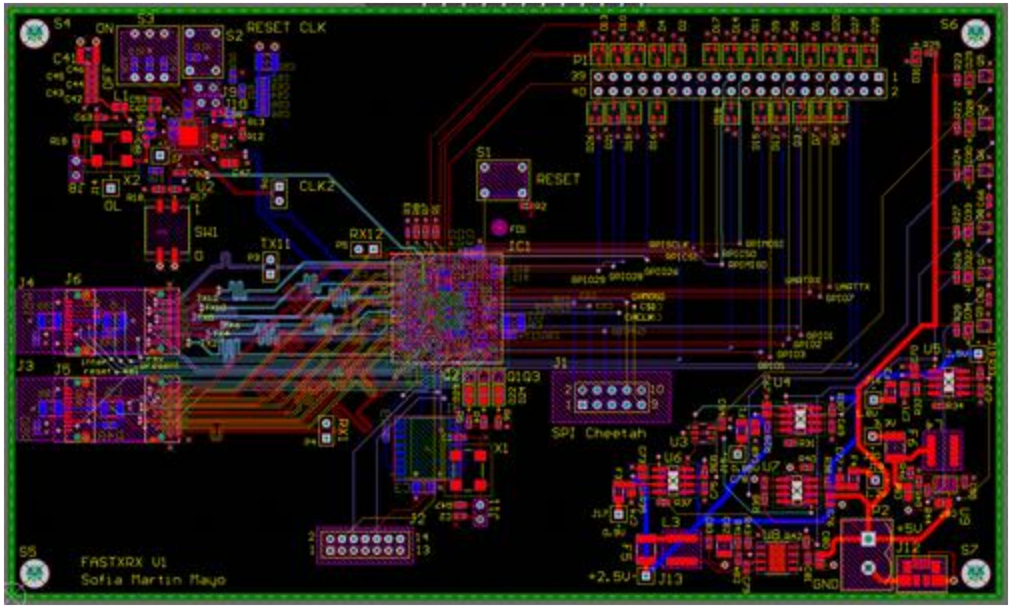


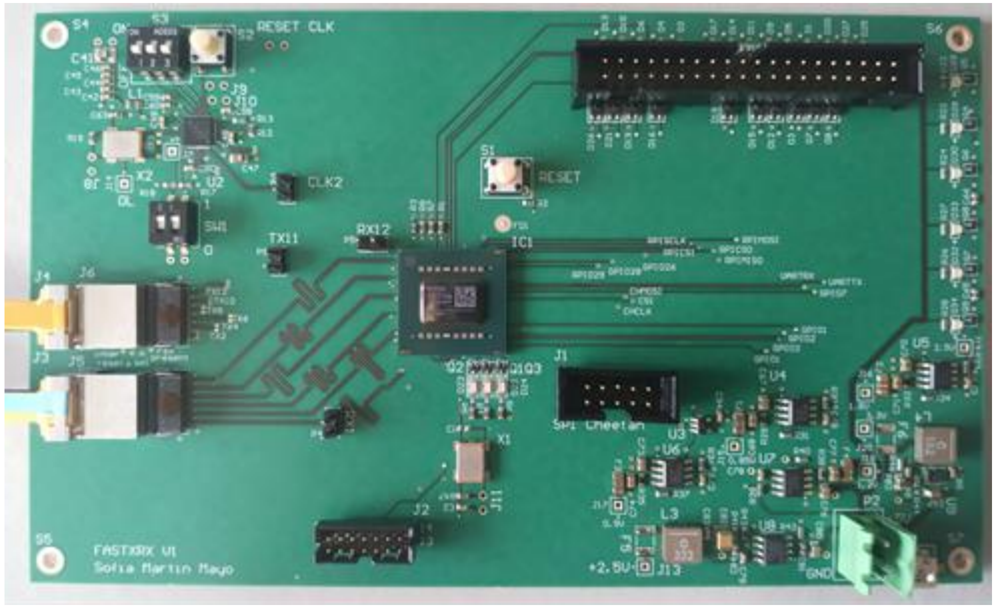
f/w simulation

ML@FPGAs for CTAO Trigger

Prototype 2: CTP high-speed testbench

- Building the testbench
  - Firefly optical transceiver + 12-layer Microwave PCB + FPGA interface.
  - Test the capabilities and quality of the companies to produce the final board.
  - Board received by the end of July. Test just started.





Team, funding & plans

- Team
  - ***LST OnSite***: 1 staff (phys), 1 predoc (phys)
  - ***CTLearn***: 1 staff (phys), 1 predoc (phys & s/w)
  - ***ML@FPGAs***: 2 staff (phys + h/w), 1 predoc (h/w), *3 predocs on 10/24 (phys + h/w + s/w)*

- Network
  - ***Spain***: CNID/COMCHA (ML@xx, OnSite Proc.); Ciemat/IFIC (ML@FPGAs), UPM (HLS)
  - ***International***: SiPM-CAM (Ciemat, UB, UniGe, UZH); CERN DRD7 (ML@FPGAs)

- Dedicated grants
  - ***Running***: *Spanish 3-year for predocs,* Spanish (PDC2023) 2-year for ML@FPGAs predocs & h/w
  - ***Requested***: Spanish 4-year for predocs & h/w, Madrid 4-year for predocs & h/w, *EU-InfraTECH 4-year (reserve list) for predocs & h/w*

Team, funding & plans

- Short-term Plans (~2026)
  - ***LST OnSite***: automatic processing to the 4 LSTs array
  - ***CTLearn***: compliance with CTAO software requirements and pipeline integration, optimize DL models for real data
  - ***ML@FPGAs***: complete 2-testbench demonstrator, deploy & benchmark simple CNNs, optimization of CNNs for FPGAs, start algorithm parallelization, start knowledge transfer

- Mid-term Plans (~2028)
  - ***LST OnSite***: integrate Onsite Pipeline in CTAO scheme
  - ***CTLearn***: CTAO offline ML-based DVR
  - ***ML@FPGAs***: build full-scale CTP prototype, deploy & benchmark optimized CNNs

Summary

- CTAO ESFRI construction started
- Recent funds → UCM-GAE involvement in ML-based R&D for CTA
- Sinergies with COMCHA teams to be pursued
- Transfer of knowledge pursued/expected from ML@FPGA activities