



TÉ DES SCIENCES

Telescope Array Observator

M. Heller on behalf of the CTAO LST Collaboration - 26/09/2024



and the second sec

Courtesy of D. Kerszberg

Real-time processing at the Large-Sized Telescopes of the Cherenkov

SMARTHEP







Outline

The Imaging Atmospheric
 Cherenkov Telescope principle

- The Cherenkov Telescope Array Observatory
- The Large-Sized Telescope of CTAO

Credit: CTAO gGmbH



Real-time processing in the Advanced SiPM Camera of LST

The project
 Data volume reduction at all stages

Real-time Analysis for CTAO: Scientific Alert Generation



Cosmic ray messengers

- Propagation is affected by
 - Intergalactic Magnetic Field
 - Intergalactic Matter/dust
 - Background Photons

p Charge particle



Neutrinos

Credit: CTAO/D. Della volpe





Imaging Atmospheric Cherenkov telescope

FACULTÉ DES SCIENCES



Detection of Cherenkov \bullet flashes require: Single photon sensitivity Nano-second time resolution









- Spatio-temporal profile of the shower image proxy to:
 - $energy \leftarrow (image intensity, impact parameter)$
 - *direction* \leftarrow (image orientation, timing) +
 - *identity of primary particle* \leftarrow (image shape, timing) +







What do we gain going for an IACT array ?

Better parameter reconstruction:
 Improved angular resolution
 Improved energy resolution
 Improved background rejection





Gamma ray astronomy

CTA Science cases

- Understanding the origin and role of relativistic cosmic particles
 - Cosmic accelerators
 - Propagation and influence of Accelerated particles
- Probing Extreme environments
 - Black hole and jets
 - Neutron stars and relativistic outflows
 - Cosmic voids
- Exploring frontiers in physics
 - Dark matter
 - Quantum gravity and Axionlike particle



















Full Sky Coverage



Southern sky



Northern sky













Cherenkov Telescope Array

	LST	MST	SST
Effective mirror area	370 m ²	88 m²	<mark>8m</mark> ²
Energy range	20GeV - 3 TeV	80 GeV - 50 TeV	1 TeV - 300 TeV
Exclusive energy range	20GeV - 150 GeV	150 GeV - 5 TeV	5 TeV - 300 TeV
#telescopes North	4	9	0
#Telescopes South	O *	14	37*
Photo-sensors	PMT	PMT	SiPM







The Large-Sized Telescope of CTAO

Stucture		
Alt-Azimuth Mount on a circular rail		
Tubular Structure in CFRP & Steel		
Full Telescope Weight	103 tons	
Maximum time for repositioning	30 s	
Optics - Parabolic Mirror		
Primary Mirror Diameter	23 m	
Focal Length	28 m	
Effective area including	370 m ²	
Camera		
Field of View	4.3 °	
Number of Pixels	1855	
Pixel size	0.1 °	
Photo Sensor	PMT	
Signal sampling rate	1 GHz	



Camera



• First prototype LST-1 inaugurated in October 2018 at La Palma, Canary Island









The Large-Sized Telescope of CTAO



M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024









Possible topologies



Triggering on Shower with LST

How does it looks like?





- Physics rate completely dominated by cosmic rays
- Operate telescopes at so-called safe threshold:
 - Intersection between:
 - trigger rate resulting from 2 x NSB level
 - trigger rate from 1.5 x protons
- This hardware threshold imposes the lowest energy achievable later at analysis stage
- Data analysis methods will improve, hardware threshold must not be the limiting factor



Night Sky Background rate

M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024









Designing the next generation of cameras for LST

How to improve performance?







• At low energies:

Images are small and faints: **Use sensors with higher sensitivity** Silicon Photo Multipliers Gammas and Protons are very lookalike **Increase image granularity** Smaller pixels





The Advanced SiPM Camera

- The proposed design should take full advantage of the SiPM characteristics
 - Gain in duty-cycle, robustness, stability, self-calibration, etc...
 - Utilise Swiss experience in using SiPM for IACT (FACT, SST-1M)
- The Advanced SiPM Camera must:
 - outperform the existing camera over the entire energy range
 - be upgradable/reprogrammable
- Baseline design:
 - Decreasing pixel size from 0.1° to 0.05°
 - 4 times more pixels !
 - ✦ Going for fully digital readout
 - Opens a lot of opertunity for real-time processing
- Many challenges to tackle:
 - High power consumption
 - High data throughput
 - High cost





M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024



4

... for more feature extraction $E_{\gamma_{PS}} = 10.89 \text{ TeV}$

Nominal LST Camera ×10³

UNIVERSITÉ

DE GENÈVE

FACULTÉ DES SCIENCES

1.0 -

0.5



-1

L. D. M. Miranda (UniGe)









M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024









Camera readout architecture



M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024



CTAO







Data Volume Reduction



Neighbor clusters -



M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024

telescopes











M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024





FADC and L1 board

• Functionalities:

UNIVERSITÉ

DE GENÈVE

FACULTÉ DES SCIENCES

- Digitize analog signals from SiPMs
- **Capture and buffer (~ 4 µs) FADC stream** •
- Perform low level trigger (digital sum)
 - 1 GHz throughput
 - ~10 ns latency
- Send to Central Trigger Processor:
 - Option A: trigger bits and data from triggered events only
 - Option B: trigger bits and high level trigger information (sum of seven pixels)
- FADC and L1 boards to be merged
- Proof-of-concept test bench setup:
 - Implementation and test for both options almost over:
 - JESD204C
 - **RoCE for Option B**
 - Implementation of low level trigger

M. Bellato, F. Marini (INFN/Padova)

M. Barcala, G. Martinez, J. Sastre (CIEMAT)







- - neighboring flowers
 - Compute sum of 7 flowers, i.e. 48 pixels and compare to threshold
 - Option A: trigger bits and high level trigger information (sum of seven pixels)
 - Option B: trigger bits and data from triggered events only











M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024





Central Trigger Processor board

- Functionalities:
 - Collect and propagate trigger bits to build coincidences
 - + Option A:
 - Collect data from events passing L1 trigger condition
 - Perform partial event building
 - Level 2 HW trigger based on full data
 - **Option B:**
 - Level 2 HW trigger based on trigger output
 - Manage trigger hardware triggers (optical link with neighbour telescopes)
- Proof-of-concept test bench setup:
 - Manufacture test PCB for optical transceivers with a FPGA tests
 - Test for porting deep learning models or DBScan to CTP FPGA



See contribution from A. Pérez

Tentative FPGA models: Kintex UltraScale KU085, KU095 or KU115



CTLearn model

Model: "CTLearn_model"		
Layer (type)	Output Shape	Param #
waveforms (InputLayer)	[(None, 30, 30, 5)]	0
SingleCNN_block (Functiona 1)	(None, 16)	1536
<pre>fc_particletype_1 (Dense)</pre>	(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)

Model synthesis

	++	+-	+.	+	+	
Name	BRAM_18K	DSP48E	FF	LUT	URAM	
DSP	++	 -	-	+ -	-	
Expression	-	-	0	2	-1	
FIFO	69	-	5814	7931	-	
Instance	3	122	17599	90345	-	
Memory		-	-	-	-	
Multiplexer	-	-	-	-	-	
Register	-	-	-	-	-	
Total	72	122	23413	98278	0	
Available	1080	1700	406256	203128	0	
Utilization (%)	++ 6	+- 7	51	48	0	





DBScan for Level 2 HW trigger

Ε

Spatial clustering algorithm:

- From: KDD-96 Proceedings. Copyright © 1996, AAAI (www.aaai.org). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu
- One "point" = one flower which digital sum is above threshold
- Determine optimal settings from kdist plot

4-dist threshold point clusters noise points

figure 4: sorted 4-dist graph for sample database 3

Time introduced in the metric to account for shower development

point iD

Initial hyper parameters of the model

- → Type of the digital sum (3n flower)
- → Threshold on the digital sum to form micro cluster (307)
- \rightarrow How do we include timing component into the metric (z = 0.05*t)
- The maximum distance between two micro clusters to form to form neighborhood (eps = <u>0.1</u>)
- → The number of samples in a neighborhood for a point to be considered as a core point (minPts = <u>15</u>).
- → Metric : euclidean distance.

L. Burmistrov (UniGe)

Two tasks: 1) Shower/NSB separation 2) Rol identification



M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024











- Challenges ahead are to run DBScan algorithm with high throughput (max. 750 kHz) and low latency (2 μs)
- Implementation study started as collaboration between University of Geneva and HEPIA (Prof. A. Upegui. Prof. Q. Berthet, T. Classen)



L. Burmistrov (UniGe)

aamma event

M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024

DBScan for Level 2 HW trigger



amma

notons







- Challenges ahead are to run DBScan algorithm with high throughput (max. 750 kHz) and low latency (2 μs)
- Implementation study started as collaboration between University of Geneva and HEPIA (Prof. A. Upegui. Prof. Q. Berthet, T. Classen)



L. Burmistrov (UniGe)

DBScan for Level 2 HW trigger





CNN for Level 2 HW level trigger

Two tasks: 1) shower/night sky background 2) gamma/hadron Random trigger: False Number of Cherenkov photons: 10526







Latency ~ 5 us CLearn

30x30 pixels x 5ns 2 CNN layers (8, 16) GlobalMaxPooling 2 dense layers (32, **2**) or (32, **3**)



Output Shape	Para
[(None, 30, 30, 5)]	0
(None, 30, 30, 8)	368
(None, 15, 15, 8)	0
(None, 15, 15, 16)	1168
(None, 7, 7, 16)	0
(None, 16)	0
	Output Shape [(None, 30, 30, 5)] (None, 30, 30, 8) (None, 15, 15, 8) (None, 15, 15, 16) (None, 7, 7, 16) (None, 16)

Total params: 1536 (6.00 KB) Trainable params: 1536 (6.00 KB) on-trainable params: 0 (0.00 Byte)

Layer (type)	Output Shape	Param
waveforms (InputLayer)	[(None, 30, 30, 5)]	0
SingleCNN_block (Functiona 1)	(None, 16)	1536
fc_particletype_1 (Dense)	(None, 32)	544
particletype (Dense)	(None, 2)	66
type (Softmax)	(None, 2)	0
======================================	KB) 8 Byte)	=====







CTAO Signal efficiency & NSB contamination - 3 classes model

- lacksquarebackground (NSB) with Cherenkov-free NSB patches.
- Histogram below shows mean classifier value of all triggered telescopes \bullet



T. Miener (UniGe)

Categorial classification task is performed including gamma/hadron separation and atmospheric shower/night sky





Signal efficiency & NSB contamination - 3 classes model CTAO

- Categorial classification task is performed including g background (NSB) with Cherenkov-free NSB patches.
- Histogram below shows mean classifier value of all triggered telescopes
- Calculation of the signal efficiency and background (N (gammaness and nsbness).



T. Miener (UniGe)

Categorial classification task is performed including gamma/hadron separation and atmospheric shower/night sky

Calculation of the signal efficiency and background (NSB) contamination for different cuts of the classifier values





CTAO Signal efficiency & NSB contamination - 3 classes model

- lacksquarebackground (NSB) with Cherenkov-free NSB patches.
- Histogram below shows mean classifier value of all triggered telescopes
- (gammaness and nsbness).



Categorial classification task is performed including gamma/hadron separation and atmospheric shower/night sky

Calculation of the signal efficiency and background (NSB) contamination for different cuts of the classifier values



c	5
ş	ς
C	-
a	ŝ,
T	2
L	,
π	Ξ
1	5
7	ζ
-	÷
1	J
Ĺ	J
Ē)
ī	Ę
1	•
۴	٦
2	2
ſ	1
•	
2	÷





Camera readout architecture





M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024





Gamma/hadron high-level software trigger system - 2 classes model

shown previously, the high-level trigger can only focus on the gamma/hadron separation.



T. Miener (UniGe)



 Assuming that the shower/NSB separation is achieved at a low-level trigger stage, e.g. using cluster algorithm such as the DB-Scan algorithm or CNN-based algorithm as



- In this particular case, we focus on Intel Altera FPGAs the PCIe400 board for LHCb
- algorithm on those FPGAs is not a "walk in the park"





C. Abellan Beteta, I. Bezshyiko (UZH)



PCIe400 synoptic

high level synthesis for machine learning





- Our finding's:
 - HIs4ml provides better results in terms of maximum achievable throughput in perspective, but currently is very limited support for Altera® boards due to deprecation of the Intel® HLS compiler. Good potential with the release of the new Intel® oneAPI backend support.
 - Initially, only 200 inferences/s were achieved on the Intel® Arria®10 PAC card for the original model for high-level triggers (6M parameters). Can we do better?
 - Increasing the clock rate to 600 MHz (standard 400 MHz) to determine the maximum achievable performance
 - ➡ 732 inferences/s.
 - Assuming an implementation with 4 instances of the inference IP in Agilex® 7
 - ➡ 2928 inferences/s.
 - The model sizes were reduced to almost two orders of magnitude:
 - By optimising the architecture based on the graph of the network, ~ 40k fps could be achieved with one instance.







• FPGA:	60000
 Prediction of throughput with AI Suite on Agilex7 performance architecture + 2 optimised 	50000 -
architectures	40000 -
 GPU: Results of throughput on Nvidia L40S GPU with Tensorflow (48 GB VRAM, 	Throughput (fps
 18 176 CUDA cores) CPU: Results of throughtput on Intel CPU with 	10000 -
Tensorflow(Intel(R) Xeon(R) Silver 4215R @ 3.20GHz)	Barvinok 3 22 AG









FACULTÉ DES SCIENCES



Camera readout architecture

RoCE protocol

Server Rack

Camera readout architecture

RoCE protocol

Server Rack

T OLLABORATION
fer
fer

Camera readout architecture FACULTÉ DES SCIENCES **RoCE protocol - Entering the NVidia realm**

 Most of the processing power is moved to GPU, simpler for porting complex algorithms at the cost of potential performance loss wrt. FPGA

Camera readout architecture

M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024

Real-time analysis

CTAO: one eye among a multisensorial environment

- All-relevant observatories in the world are interconnected via alert systems
 - + LST and to further extent, CTAO must be able to receive and emit alerts
 - Extremely important for transient phenomenon, e.g. Gamma Ray Burst
 - Upon reception of a high-priority alert (criteria depends on observatory and moment), observation schedule is altered • Some alerts do not provide precise direction, CTAO might
 - improve there
 - During survey's or observation, some transient acitivity might be reported by CTAO itself
 - Some sources may flare while being observed, need to inform other observatory about the increase in activity.
 - Need of a real-time analysis pipeline

AstroColibri

https://doi.org/10.3390/galaxies11010022

LST Analysis pipeline

Classical approach

Raw event Calibrated event

Credit: A. Moralejo, LST Analysis school

Integrated charge (p.e.) Peak time (ns)

Regression for direction and energy reconstruction, and γ/h classification performed with dedicated random forests

Real-time analysis

The Science Alerte Generation system

• Aim is to provide real time feedback on the telescope observation (for transient and variable source observations)

- Run optimized version of the off-line analysis software, coded in C++
- In total, 66 seconds are needed to produce a DL3 from 20k R0 events (scales linearly with number of events)
 - offline analysis 30 ms/event
 - RTA ~0.5 ms/event

- one new file every 20k events
- 4 streams x 1 cores used
- rate per core ~ 2000Hz

Real-time analysis

Performance

- Crab field sensitivity ~ Istchain x 4
- Mkn 421 sensitivity ~ lstchain x 2
- RTA results seems highly affected by NSB

S. Caroff, P. Aubert, V. Pollet, T. Vuillaume (LAPP)

 Performance mostly affected by nonoptimized image cleaning with respect to classical analysis

M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024

Camera readout architecture

M. Heller | Real-time processing for LST | EDGE ML School 26/09/2024

"Smart sensors"

Use of AI ASICs for sensor real-time processing

- Goal: EdgeML accelerator ASIC for targeting real-time waveform analysis
 - Temporal signal analysis (1D CNN, ...)
 - (Data reduction, e.g. auto-encoder for denoising)
- CSEM edgeML technology:
 - CSEM Portfolio of different IP blocks to build system-on-chip

Computational complexity

"Smart sensors"

Developing single photo-electron extraction

- Identify single photons out of uncalibrated waveforms
- Output is a quantity of photons per unit of time
- Baseline algorithm already developed for CTA based on 1D CNN ("only" 28k parameters)

- Study feasibility to run realtime inference
 - Risk mitigation scenario The algorithms runs on preselected waveform (after low level trigger implement in FPGA)
- Project updated in collaboration with HEPIA Prof. A. Upegui, Y. Perin

"Smart sensors"

Digital Photon Conter (DiPC)

SPAD

- aqua SPAD #31
- PDP = [10, 35]% @ [320, 500]nm @ 3.3V excess bias

6 × 6

CHANNEL

Comparator

- The resistors R₁ and R₂ convert currents I₁ and I₂ into voltages V₁ and V₂
- The buffers B₁ and B₂ digitize voltages V₁ and V₂ into signals D₁ and D₂

 V_2

 R_2

 V_1

 R_1

E.Bernasconi, E. Charbon (EPFL), A. Biland (ETHZ)

PIXEL

• 40 × 40 = 1'600 channels

Architecture

- A binary tree adder implements the summing
- A 4 bit bus outputs the result every 4 ns (L0 trigger)
- A 12 bit bus outputs the result on demand (L1 trigger)

Conclusions and Perspectives

- Gamma-ray astronomy already relies on real-time processing due the to nature of observations
 - Constantly looking at the sky, awaiting for (non-bunched) Cherenkov flashes
 - Variable observing conditions
- Sensitivity of instruments are improving, but major steps in gamma-ray astronomy in the future will also heavily rely on real-time processing: FADC cost and power consumption going down, cameras will unevitably move to fully digital readout opening new opportunities:
 - - Sensor signal treatment
 - More sofisticated low level trigger algorithm to fight against the NSB wall \bigcirc Robust and efficient high level trigger to to particle classification Fast and performing real-time analysis pipelines for better reaction to alert or emission \bigcirc
- Stay tuned, more telescopes are coming and CTAO will soon deliver science as an array!

Credit: CTAO gGmbH

Minister Stations alastic

