Institute for theoretical physics IV
Chair for Plasma Astroparticle Physics

DATA VOLUME REDUCTION FOR THE LST PROTOTYPE

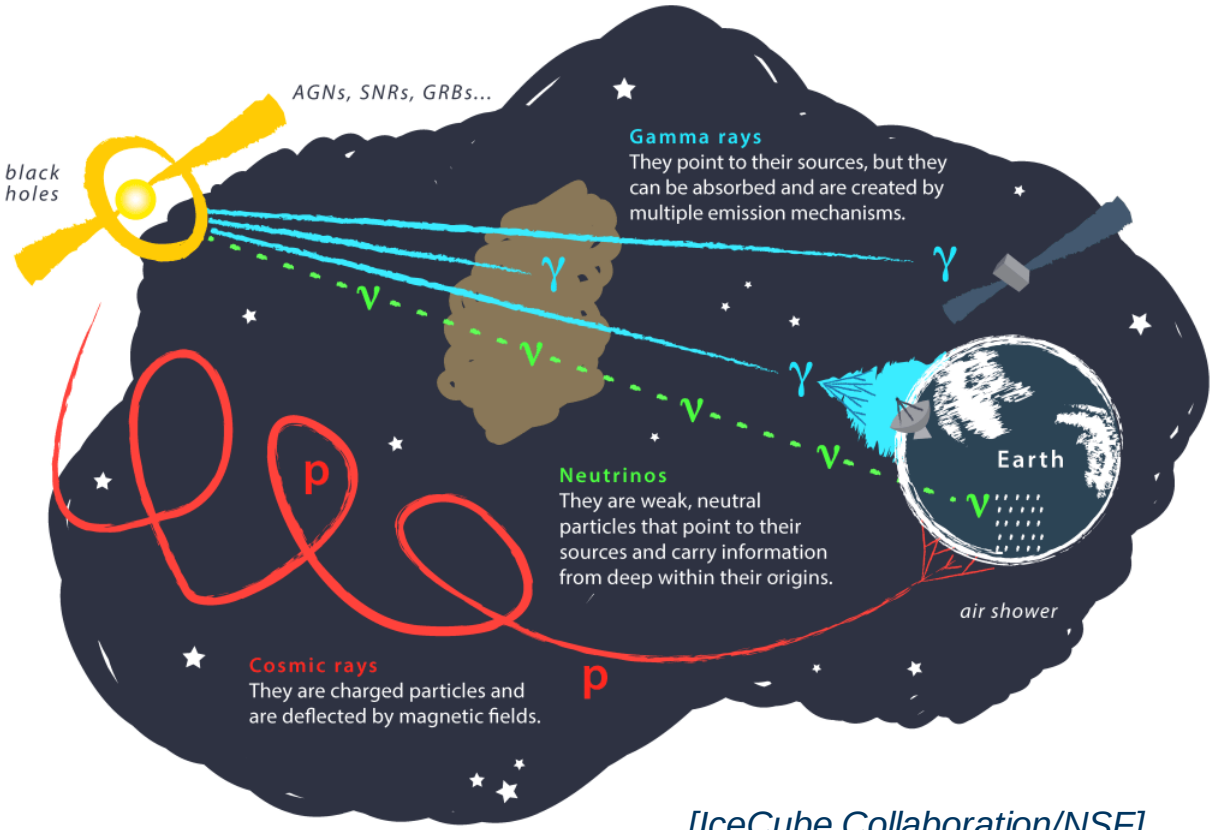
17.12.2021

Jonas Hackfeld

RUHR-UNIVERSITÄT BOCHUM

Table of Contents

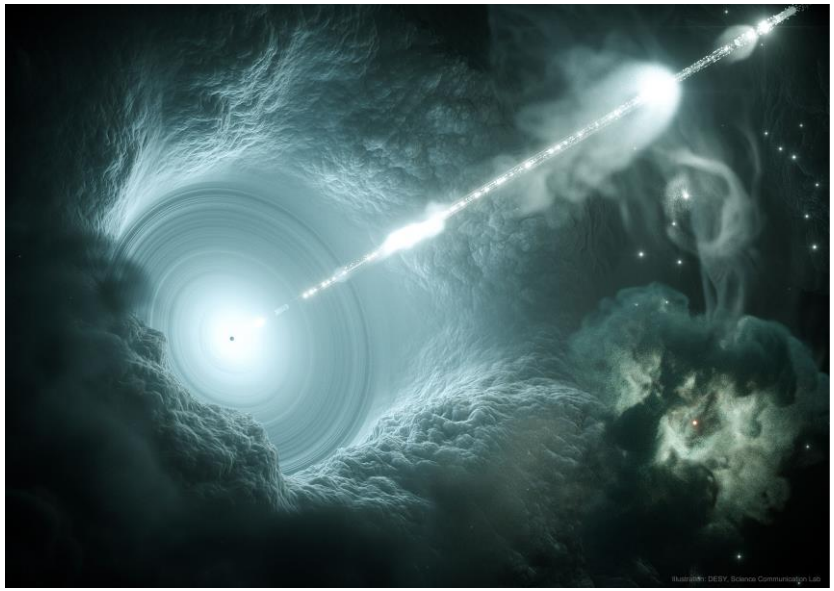
Multi-Messenger Astronomy



[IceCube Collaboration/NSF]

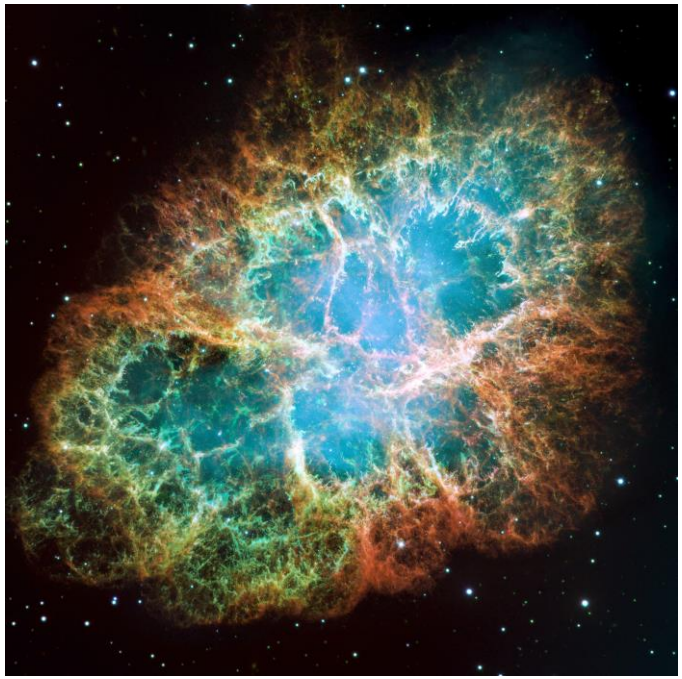
Sources in the Universe

Active Galactic Nuclei, AGNs



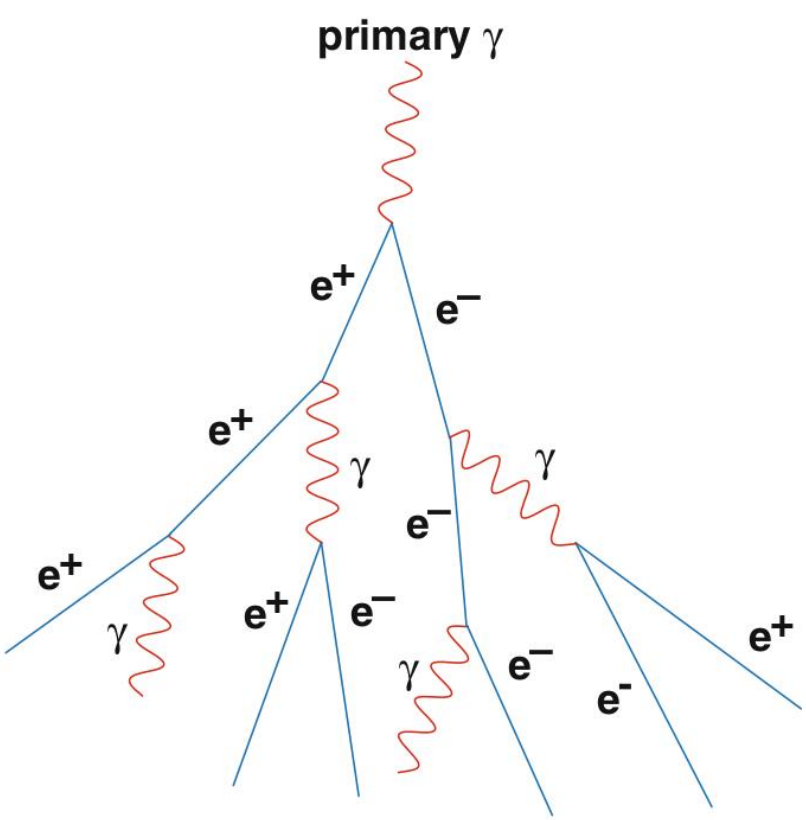
[Artistic impression of an AGN, APOD 2018-07-16]

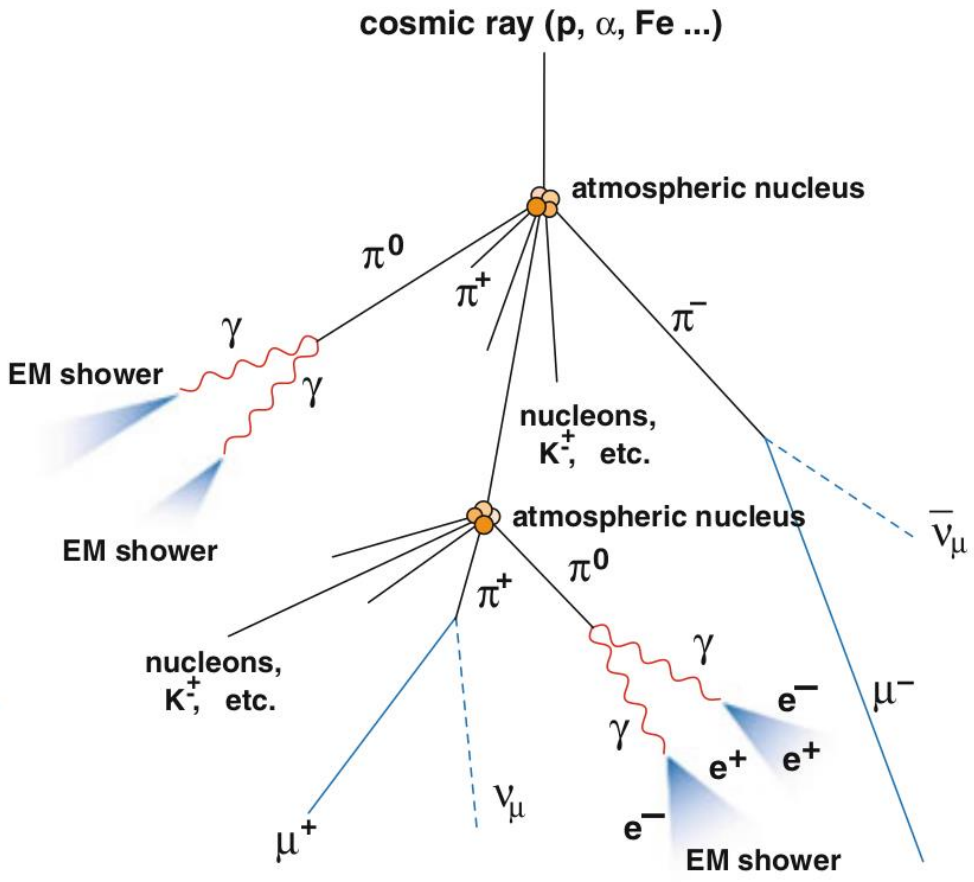
Supernova Remnants, SNRs



[HST image of the Crab Nebula]

Extensive Air Showers

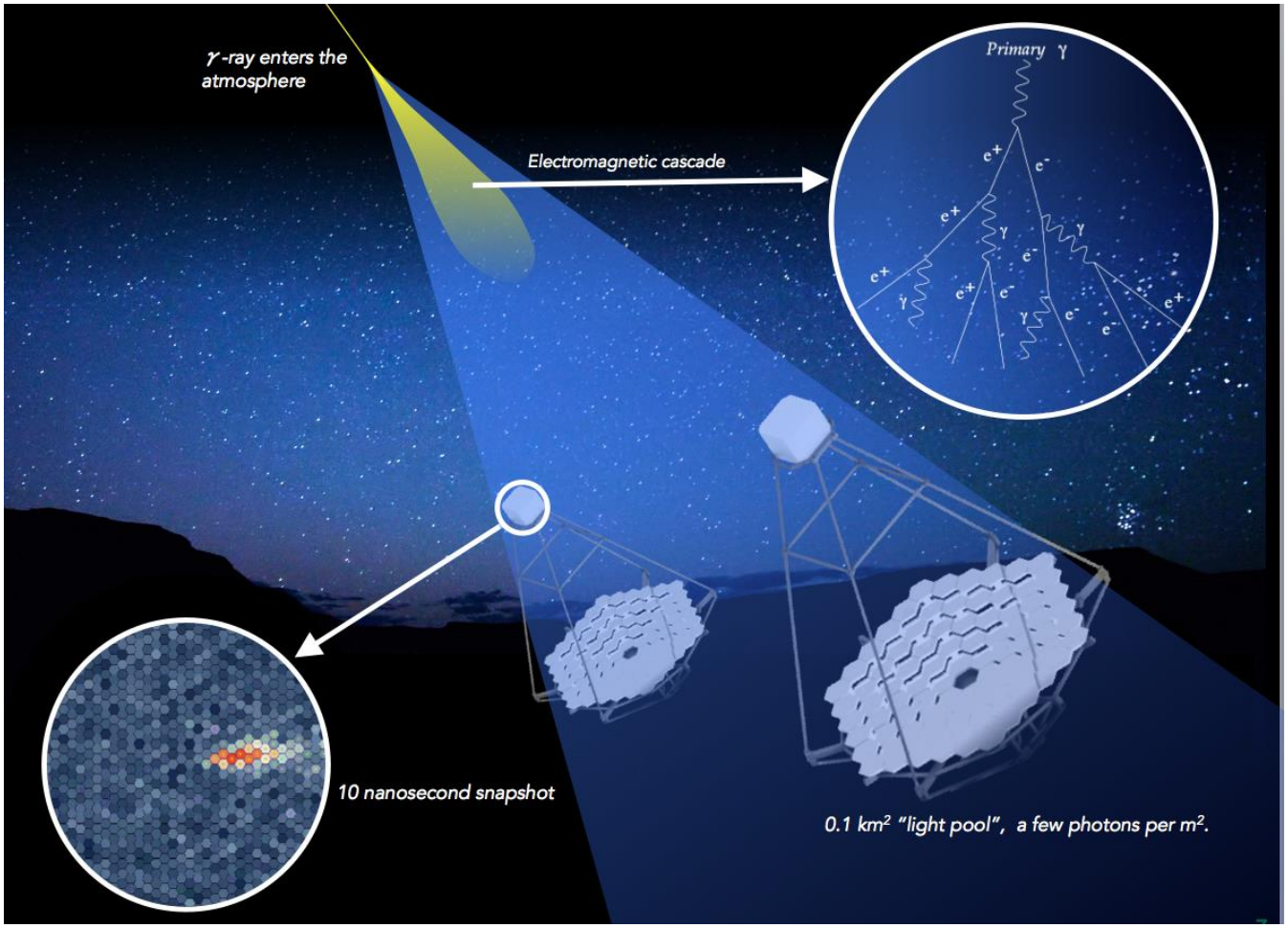




[M.Nöthe]

RUHR UNIVERSITÄT BOCHUM RUB

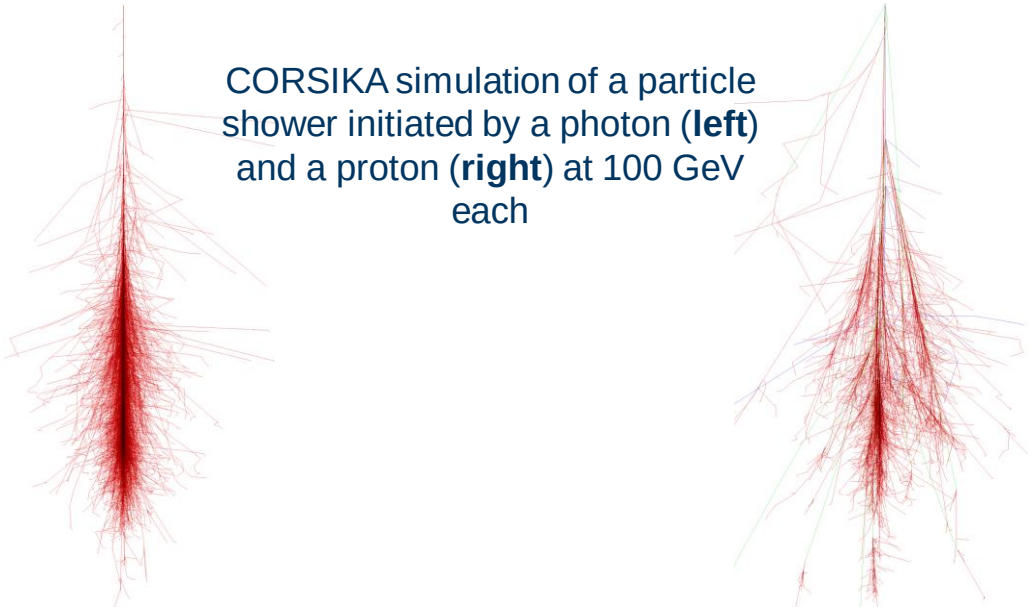
IACTs

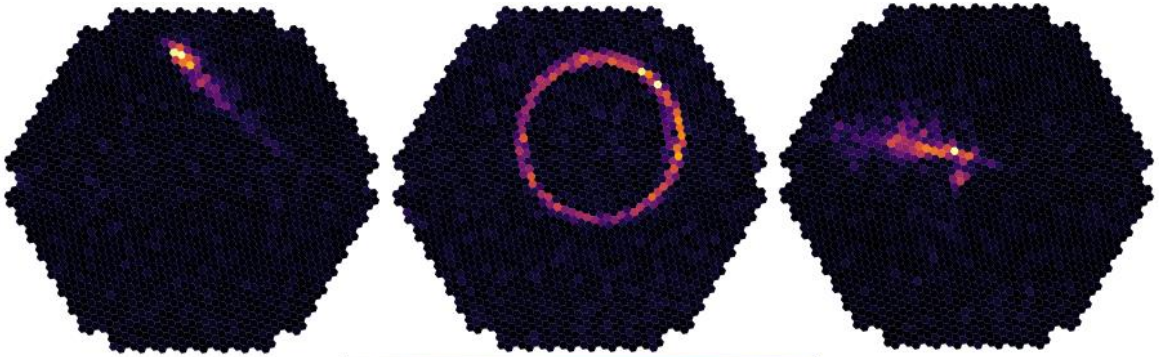


[CTA Observatory]

IACTs

CORSIKA simulation of a particle shower initiated by a photon (**left**) and a proton (**right**) at 100 GeV each





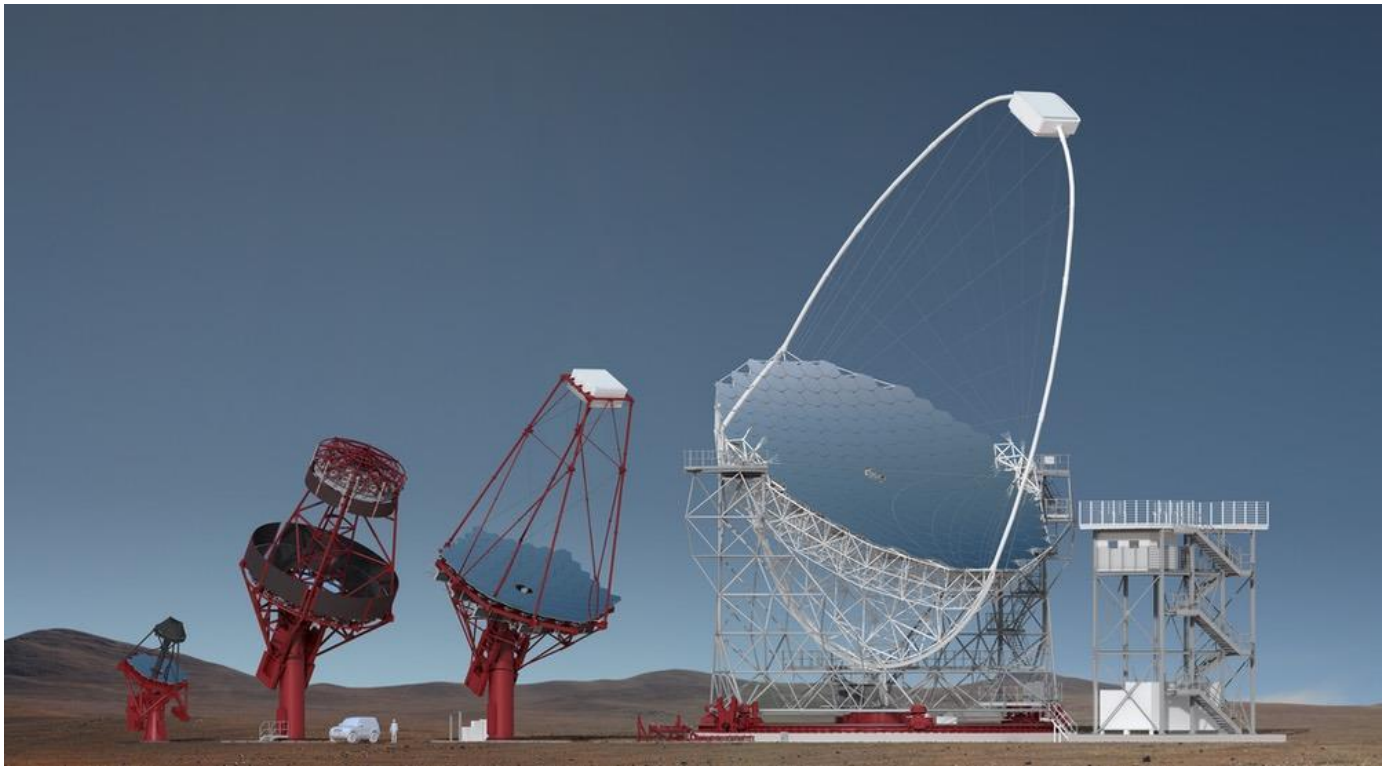
Photon

Muon

Proton

RUHR UNIVERSITÄT BOCHUM RUB

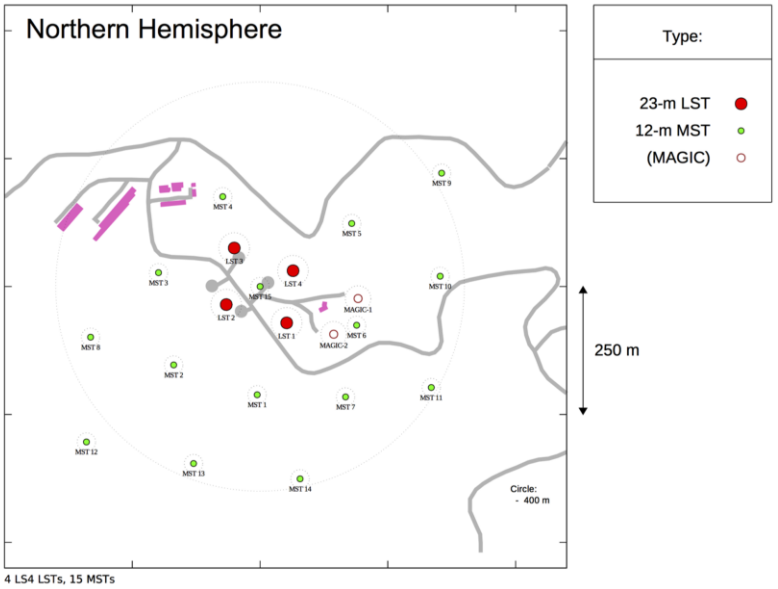
Cherenkov Telescope Array (CTA)



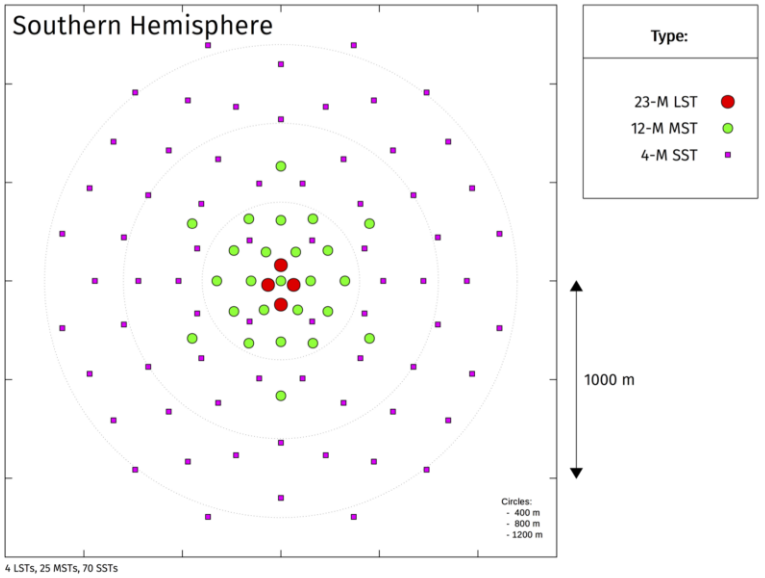
Telescope	SST	MST	LST
Energy range	5-300 TeV	150 GeV – 5 TeV	20-150 GeV
Field of view	10.5°	7.5° - 7.7°	4.3°
Reflector diameter	4.3 m	11.5 m	23 m

Cherenkov Telescope Array (CTA)

Omega configuration:



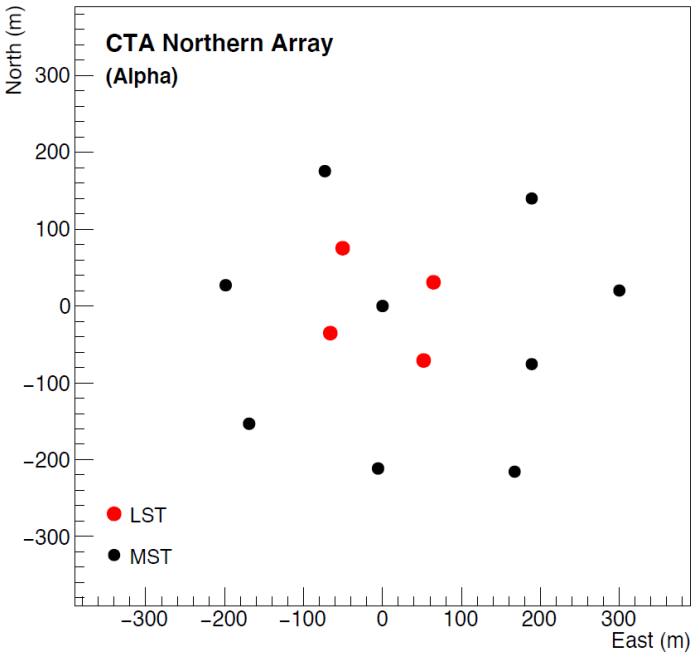
La Palma, Spain



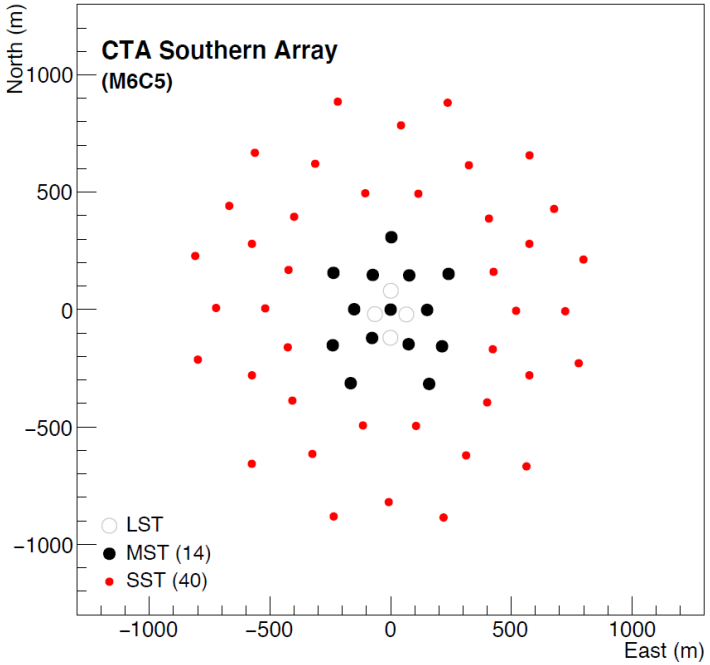
Atacama Desert, Chile

Cherenkov Telescope Array (CTA)

Alpha configuration:



La Palma, Spain



Atacama Desert, Chile

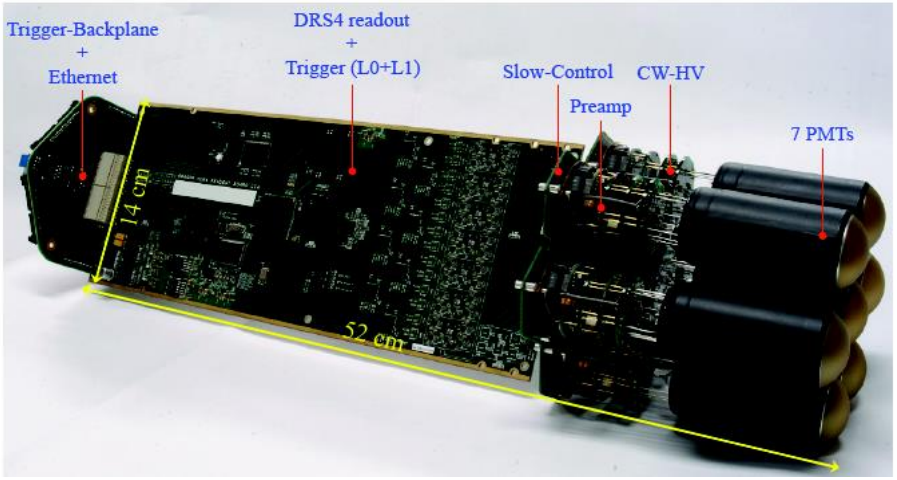
[Orel Gueta, ICRC21]

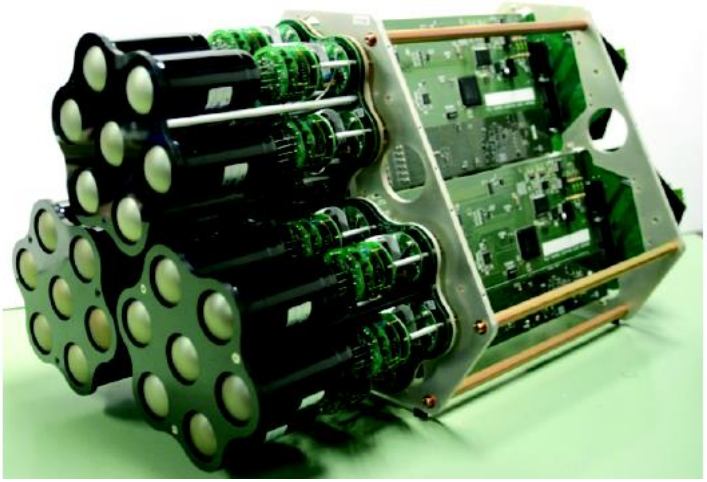
RUHR UNIVERSITÄT BOCHUM RUB

LST Prototype

[CTA Observatory flickr]

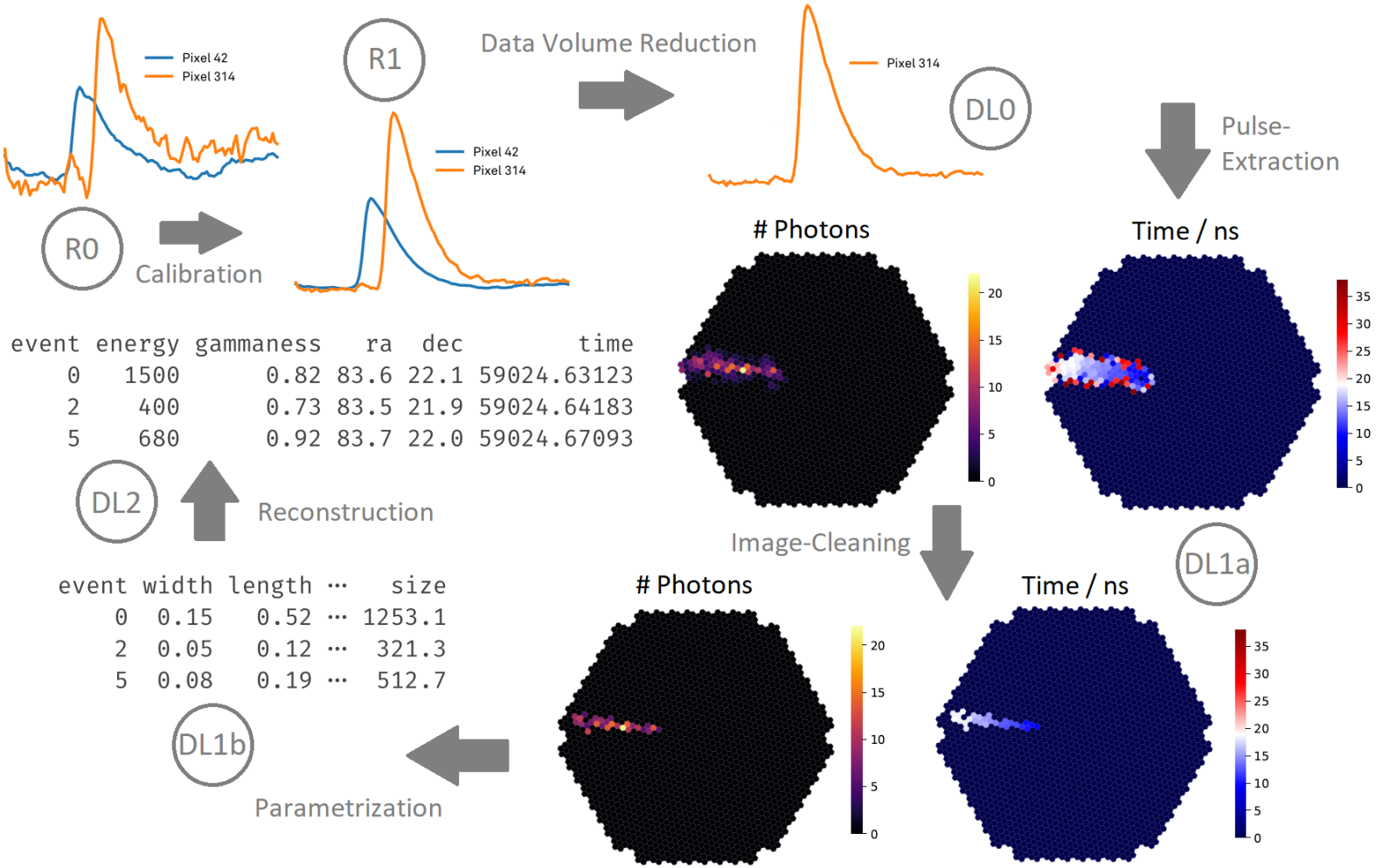
LST Prototype - Readout





Waveform

LST Prototype – Data Analysis Steps



RUHR UNIVERSITÄT BOCHUM RUB

LST Prototype - Software

- **ctapipe**: Main processing pipeline for low level CTA data – Python based framework and hosted as an open-source project on GitHub
- **ctapipe-io-lst**: Plugin for ctapipe, which is able to read LST zfits files and calibrate them to R1-level as needed for the ctapipe tools
- **aict-tools**: Used to apply machine learning tasks for event reconstruction. It itself uses scikit-learn.
- **pyirf**: Main features include calculation of the Instrument Response Function (IRF) and sensitivity estimation of the array. Utilized in this thesis for cut optimization and event weighting

➢ **MC's**: South-pointing PROD5 simulations of the northern array with 20° zenith angle and 80% mirror reflectivity

Repository	Version
ctapipe	0.10.5 – 2021-03-04
ctapipe-io-lst	0.10.0 – 2021-05-05
aict-tools	0.27.1 – 2021-05-05
pyirf	0.5.0 – 2021-05-11

Data Volume Reduction

- Permanent storage of hundreds of PByte/year over a planned 30-year lifetime will be financially unsustainable.

Lossless:

- Data compression

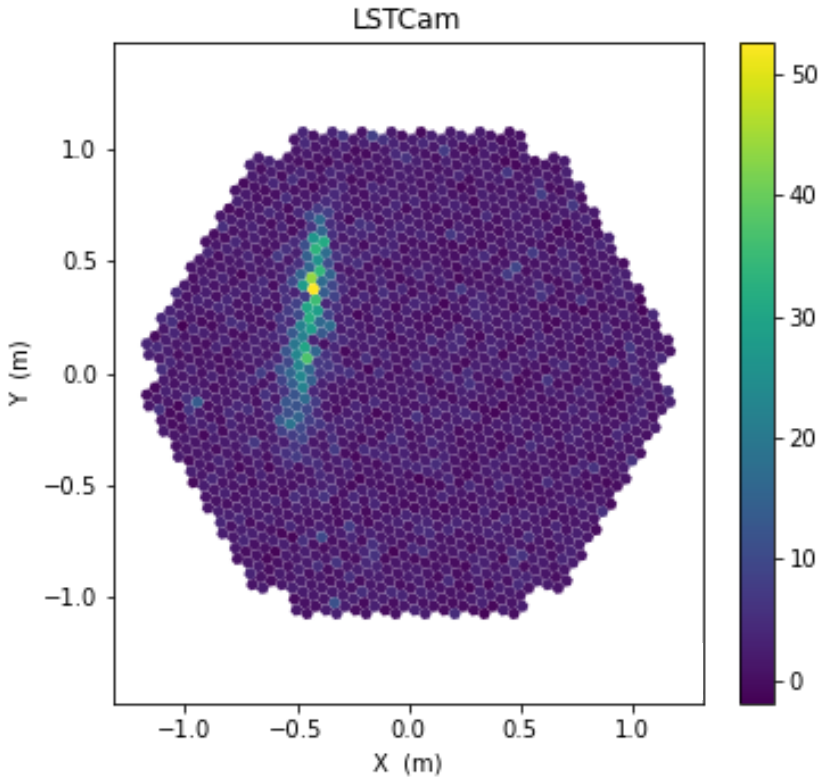
Lossy:

- Low/high gain selection
- Suppression of hadronic events
- Waveform parametrization
- Signal-pixel selection this thesis

Pixel selection

Approaches:

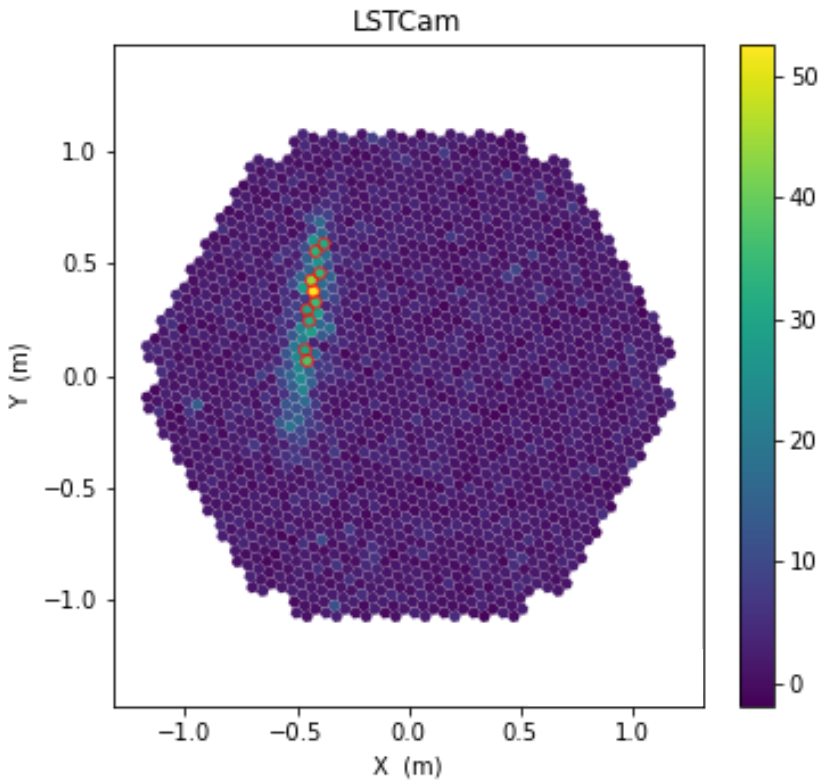
- Wavelets cleaning
- Deep-Learning
- Tailcuts cleaning



Pixel selection

Approaches:

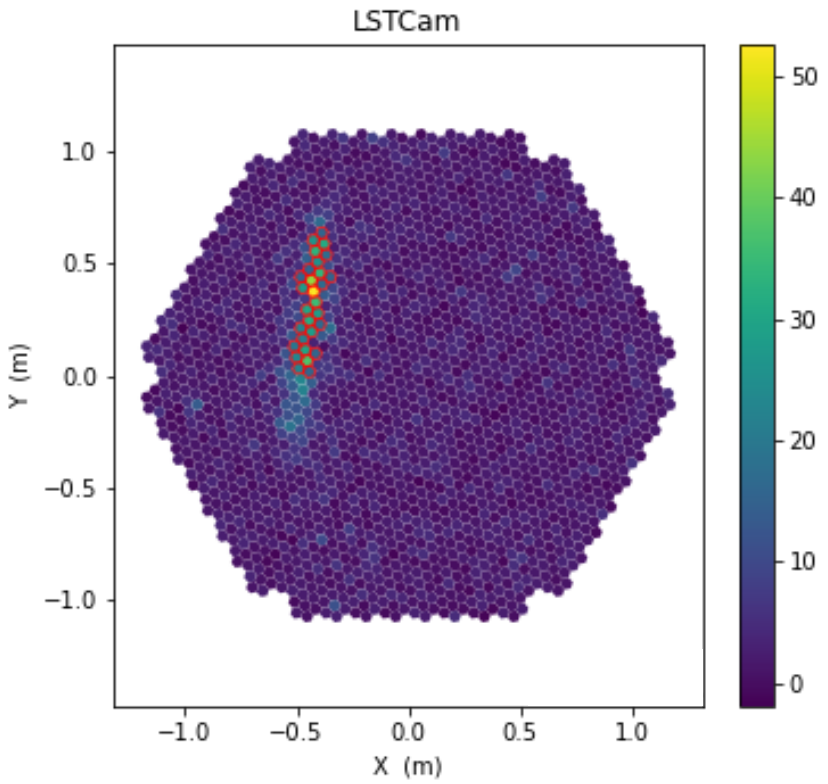
- Wavelets cleaning
- Deep-Learning
- Tailcuts cleaning



Pixel selection

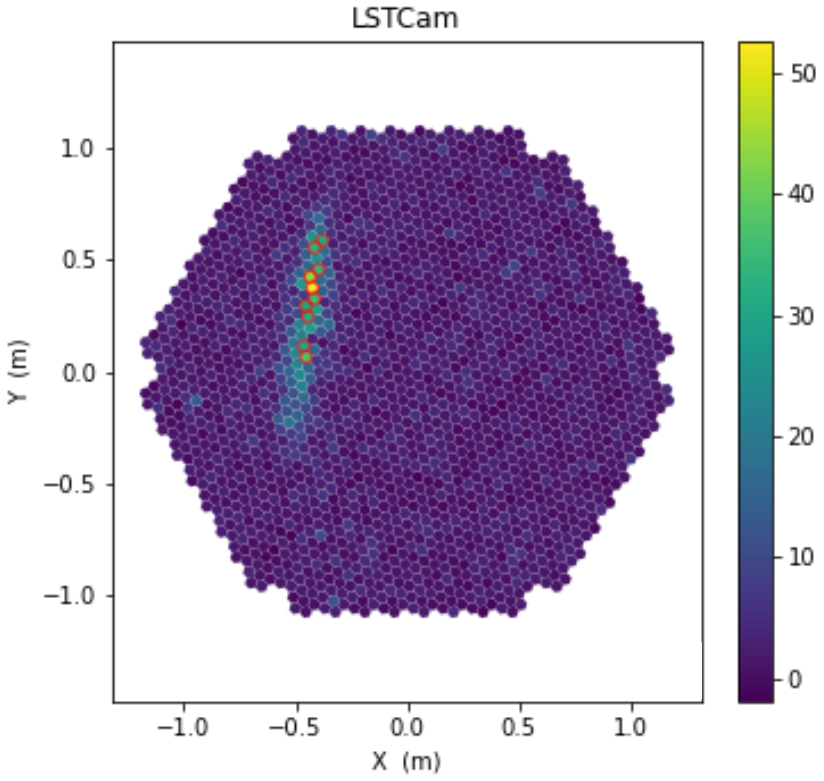
Approaches:

- Wavelets cleaning
- Deep-Learning
- Tailcuts cleaning



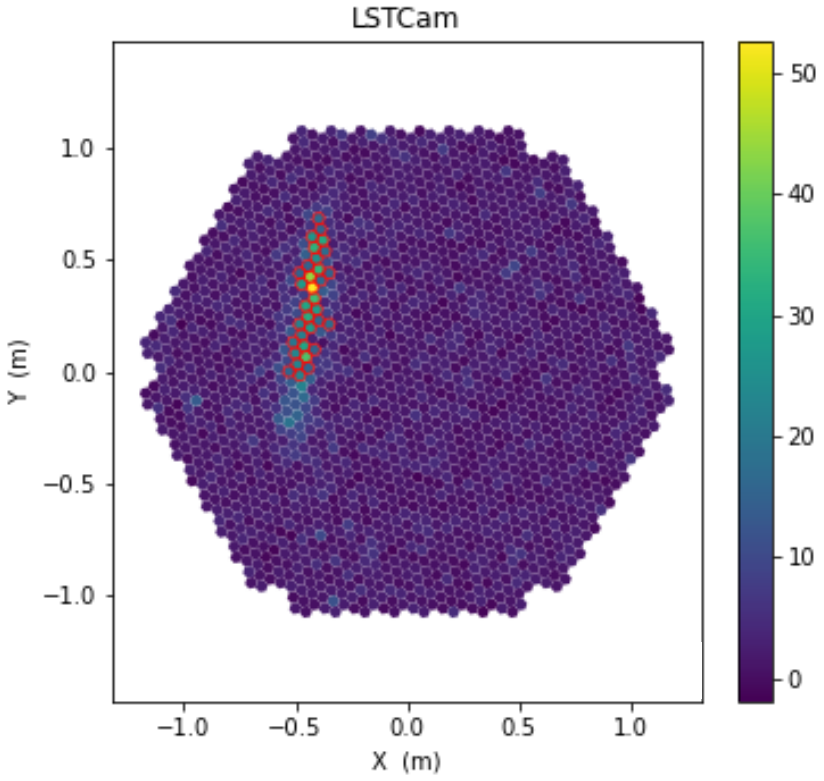
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning



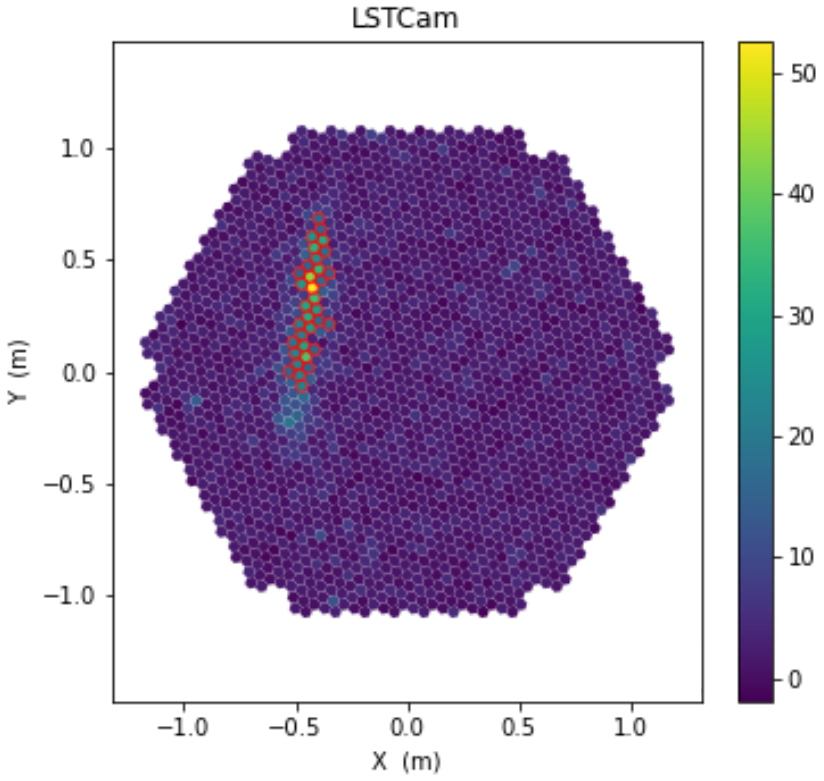
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning
2) Iteratively adding neighbouring pixels above boundary threshold



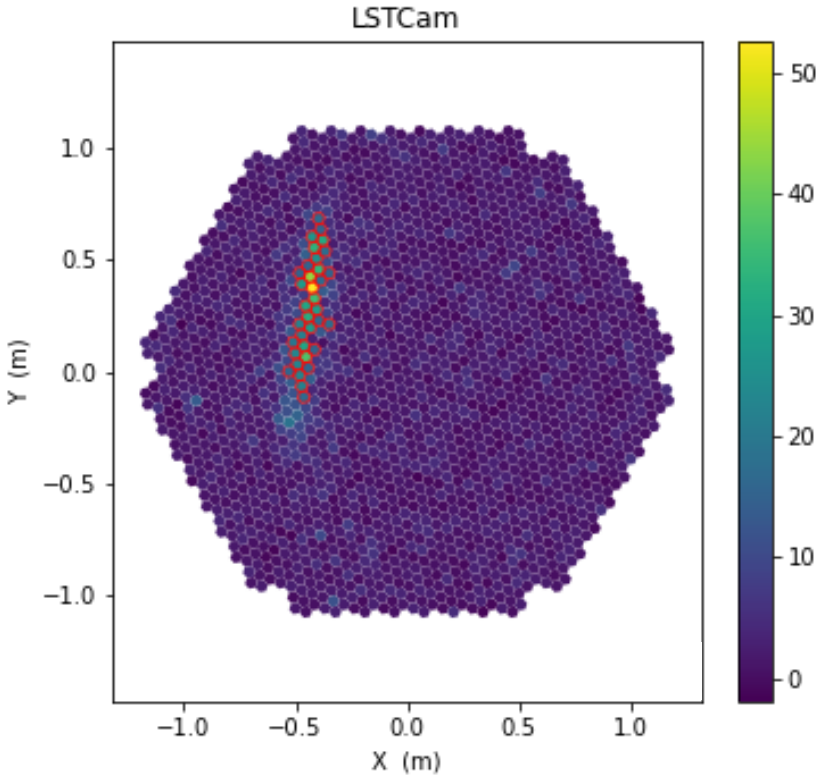
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning
2) Iteratively adding neighbouring pixels above boundary threshold



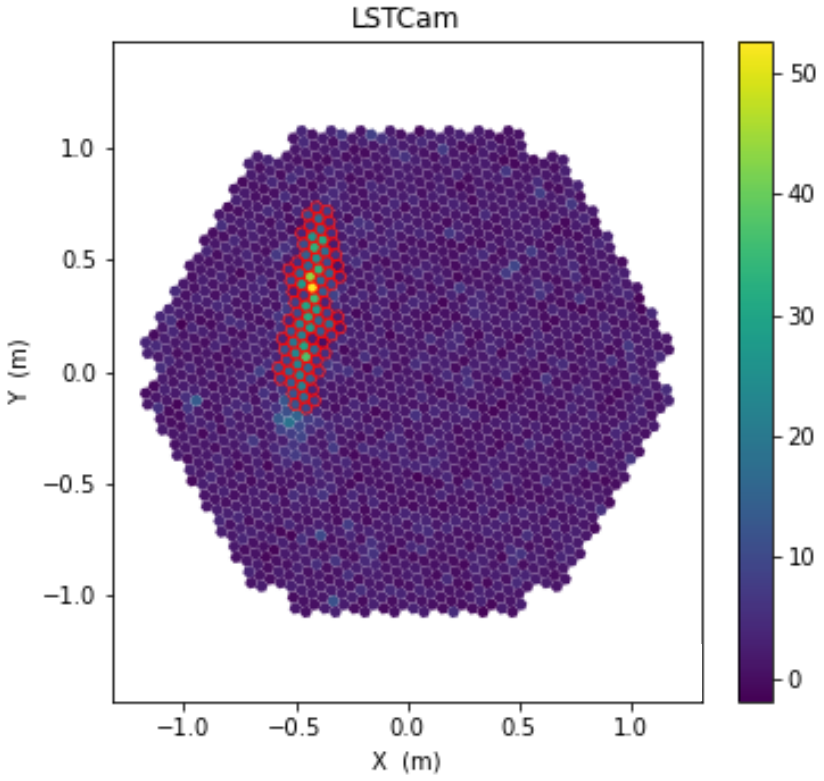
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning
2) Iteratively adding neighbouring pixels above boundary threshold



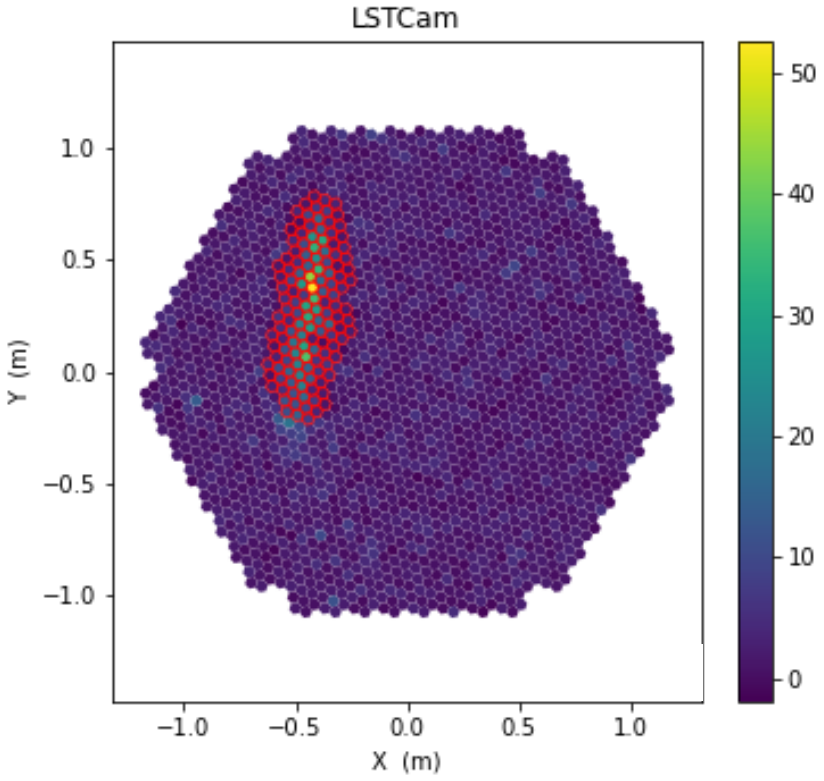
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning
2) Iteratively adding neighbouring pixels above boundary threshold
3) At the end: Adding all neighbouring pixels



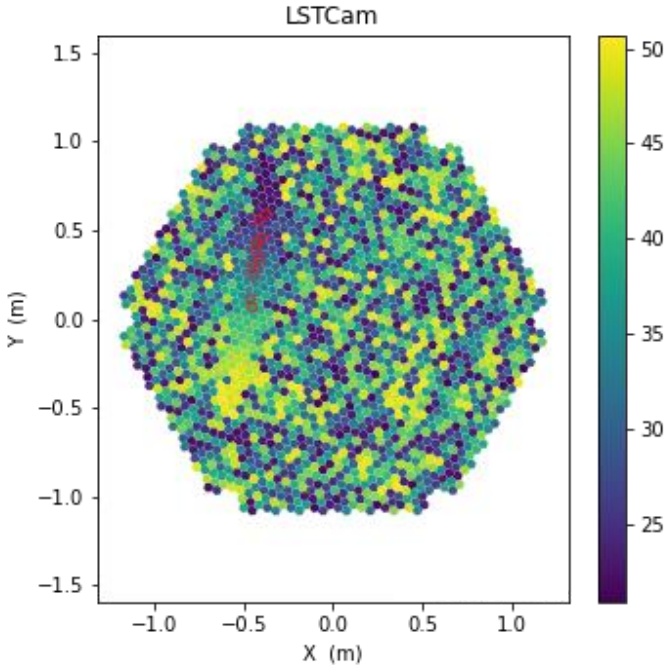
TailCutsDataVolumeReducer

1) Normal Tailcuts cleaning
2) Iteratively adding neighbouring pixels above boundary threshold
3) At the end: Adding all neighbouring pixels

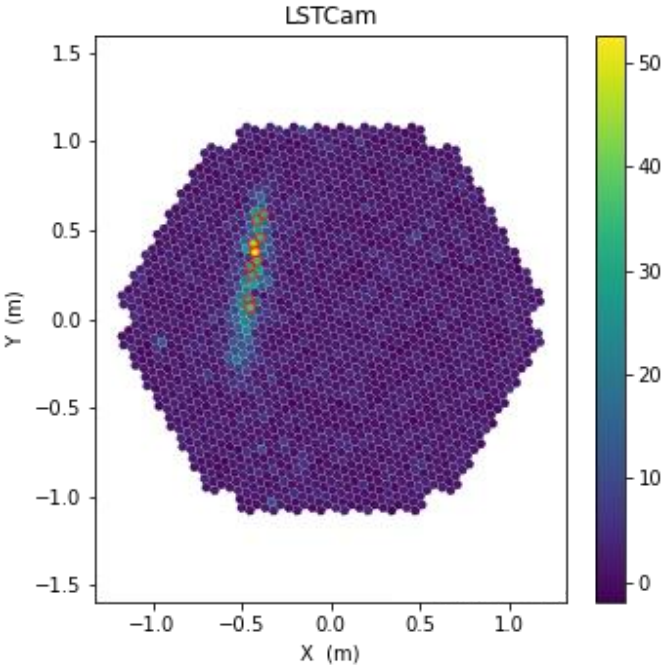


ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons



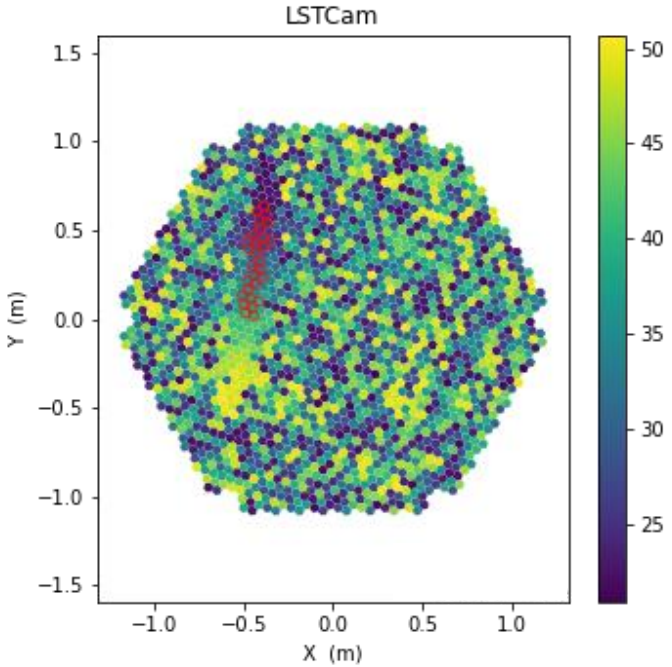
Time/ns



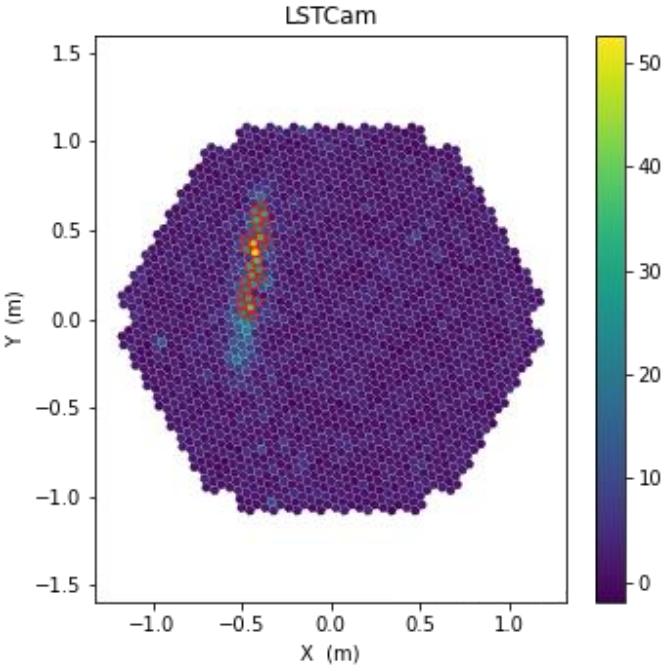
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons



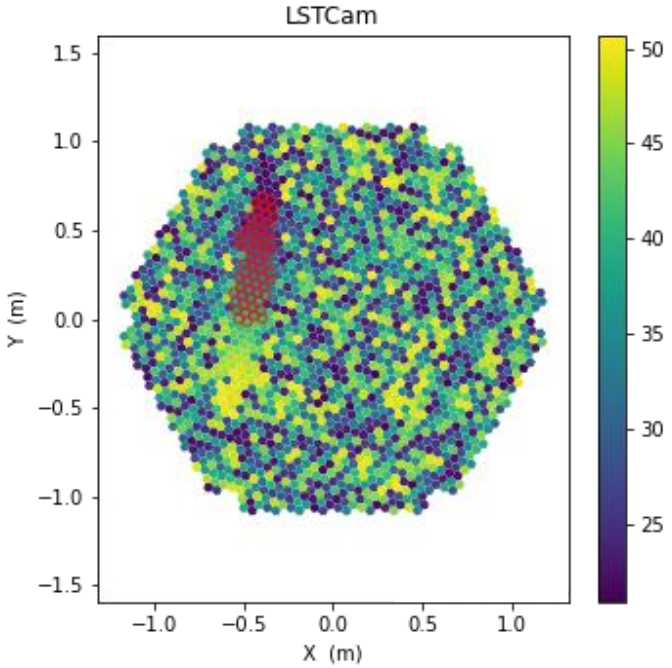
Time/ns



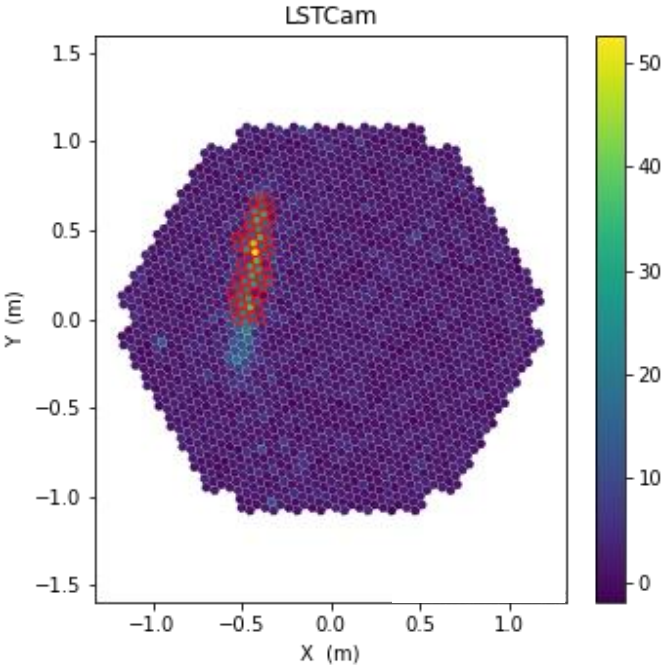
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



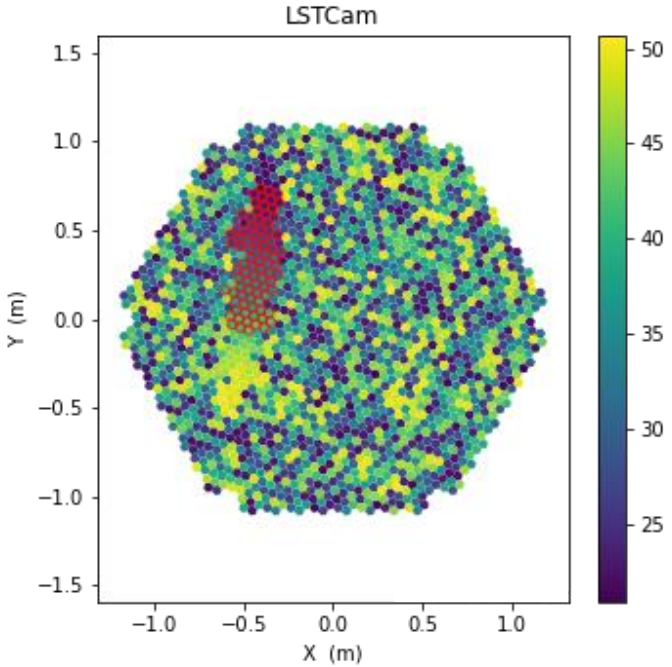
Time/ns



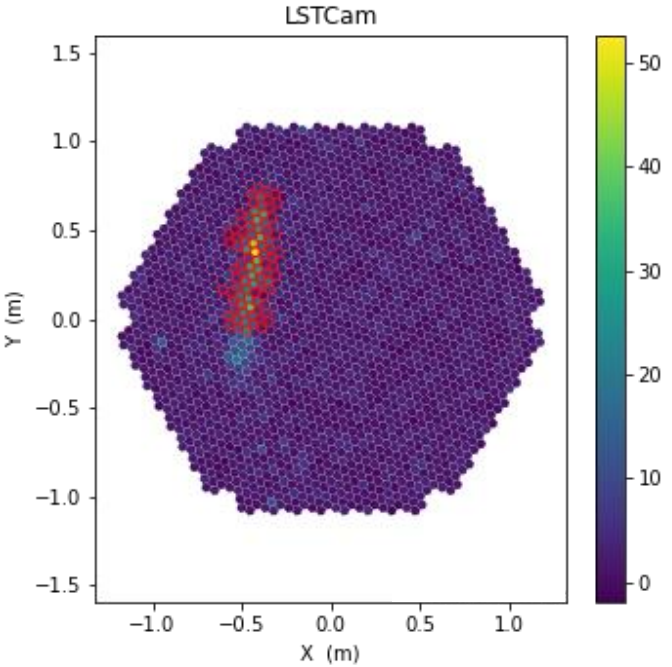
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



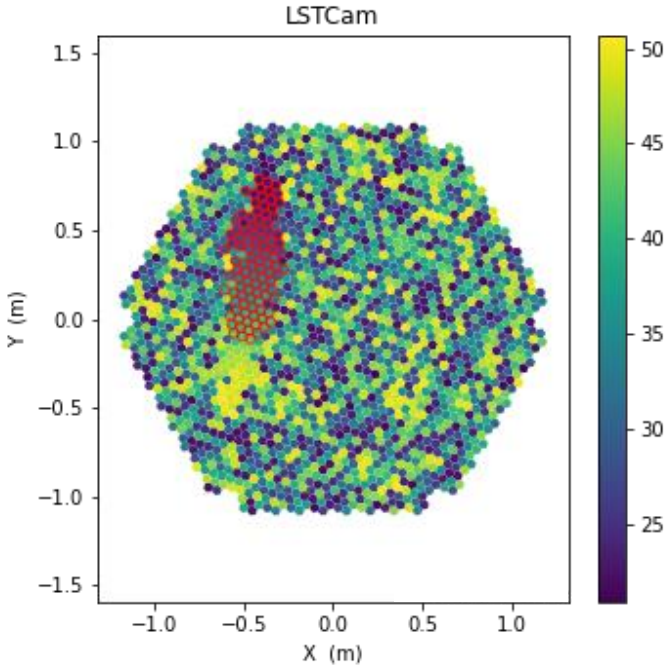
Time/ns



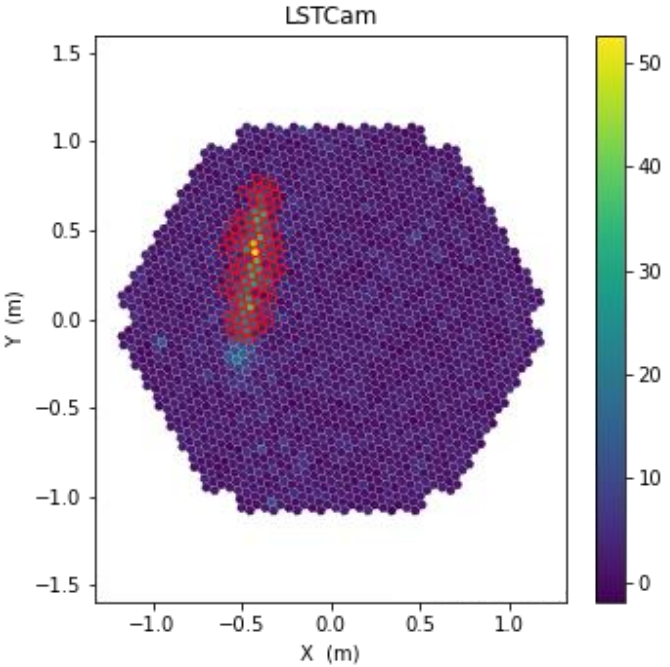
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



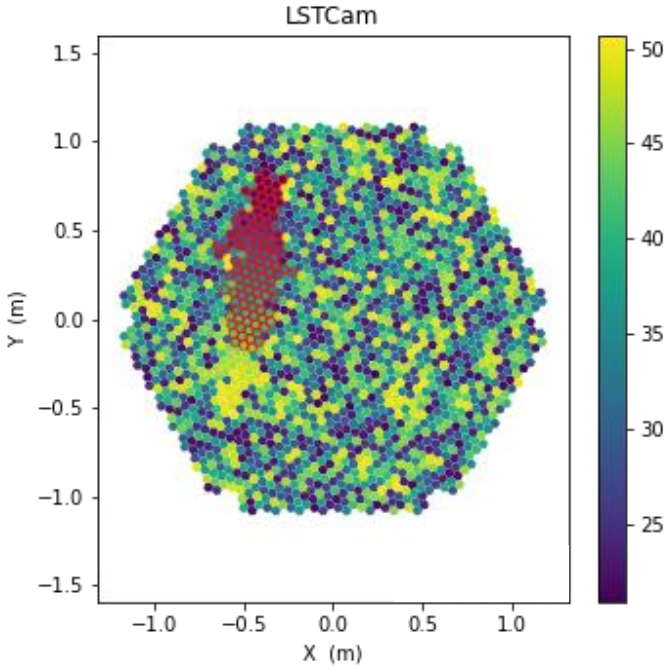
Time/ns



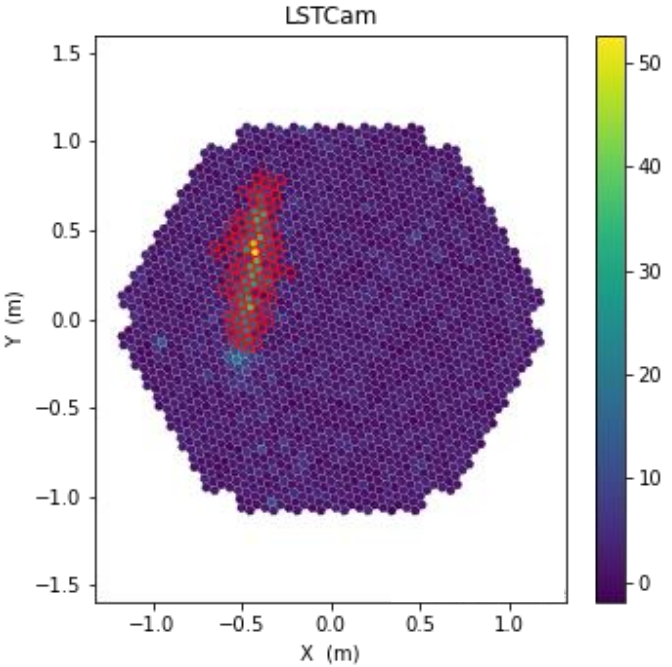
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



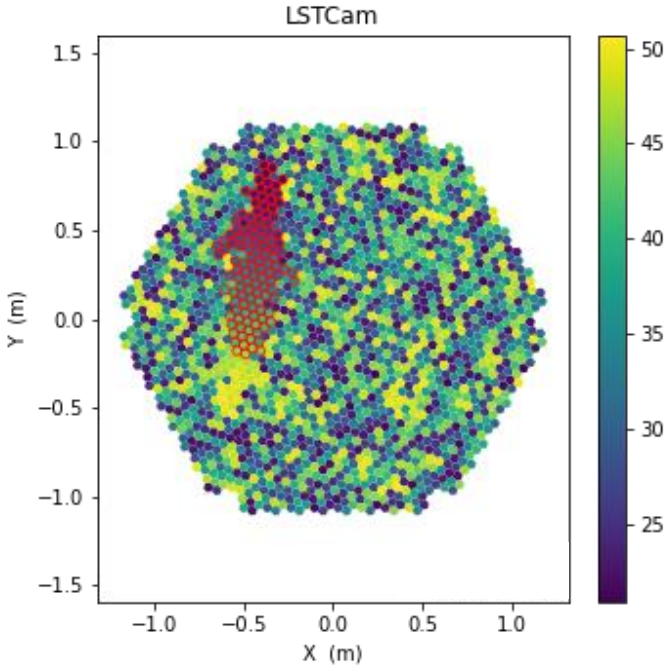
Time/ns



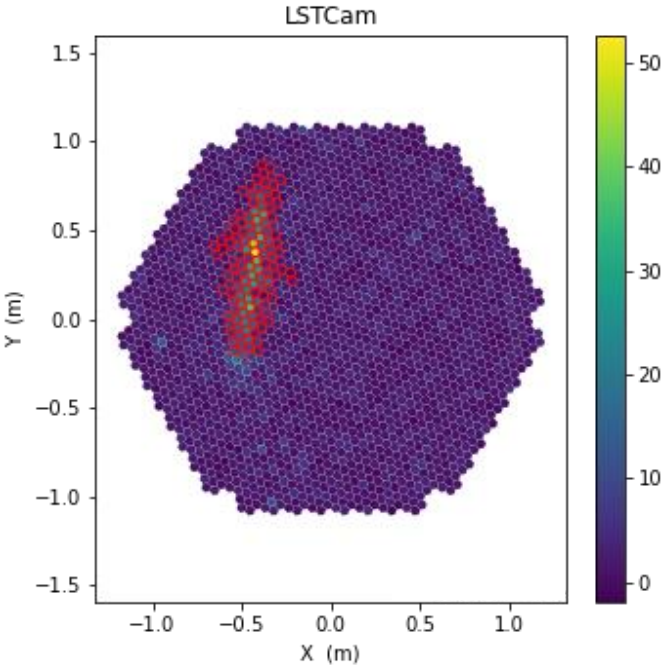
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



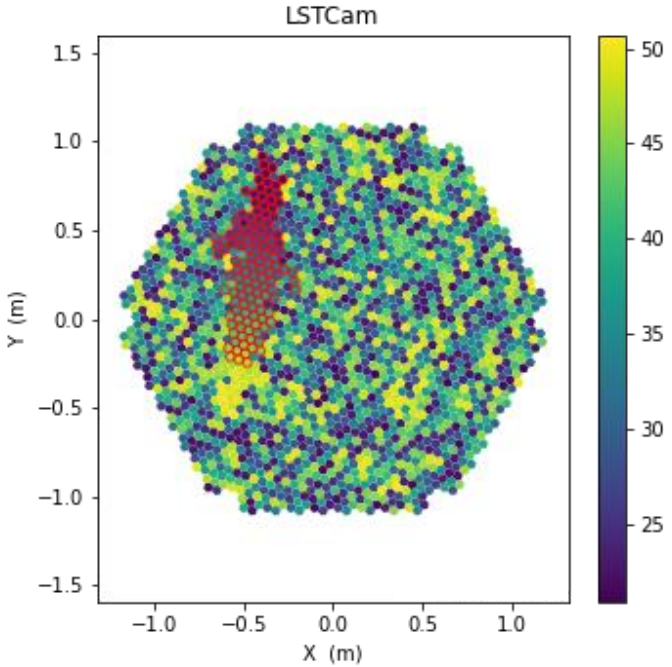
Time/ns



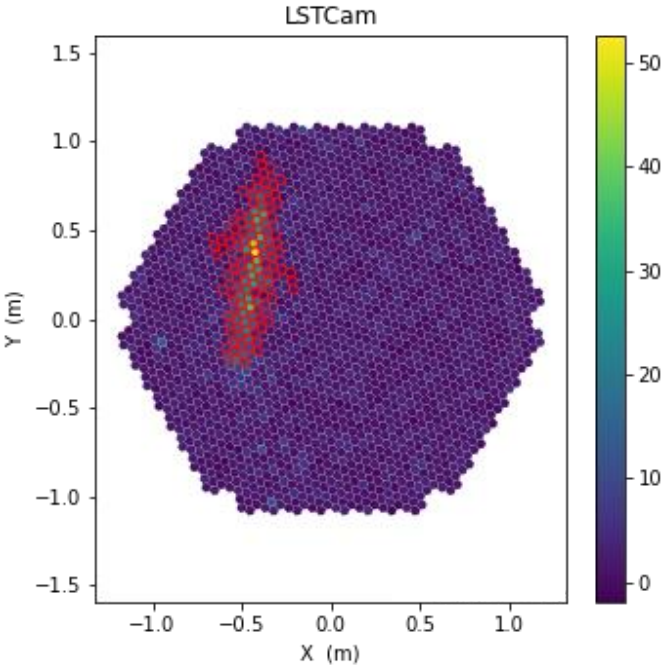
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



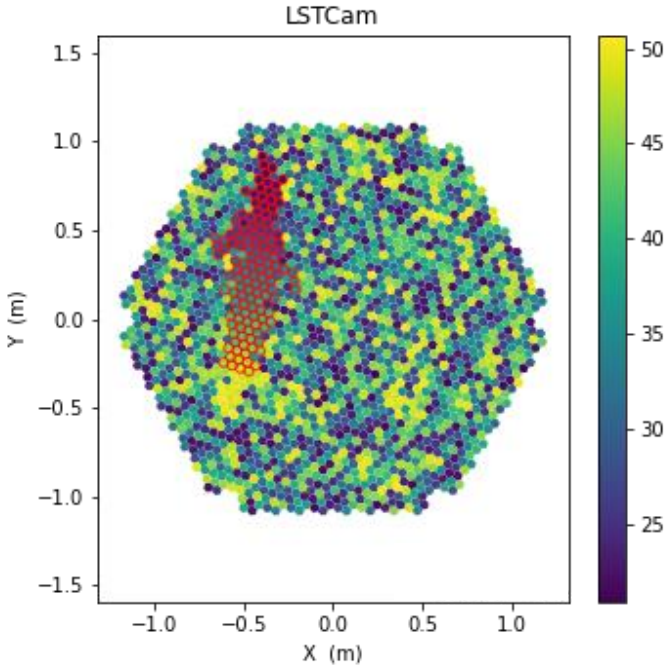
Time/ns



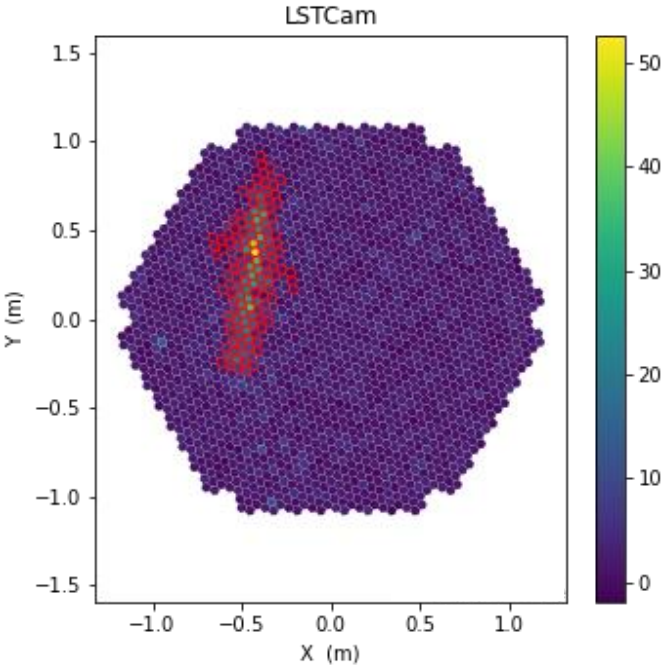
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



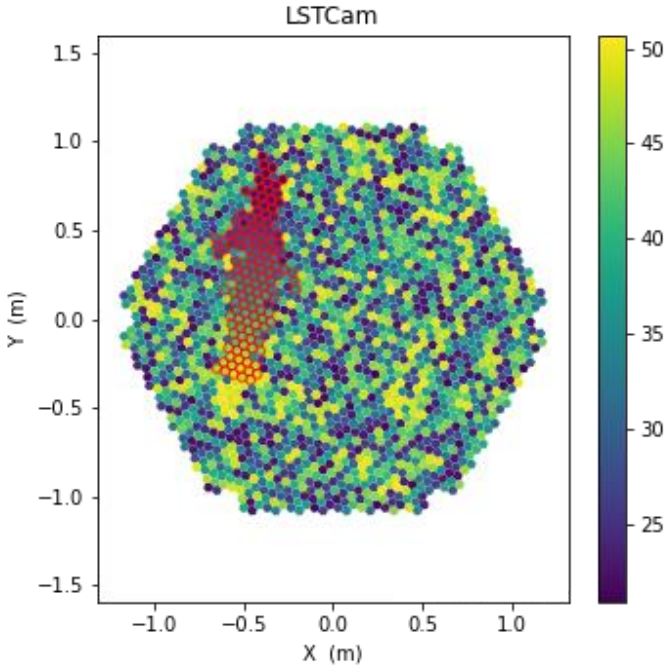
Time/ns



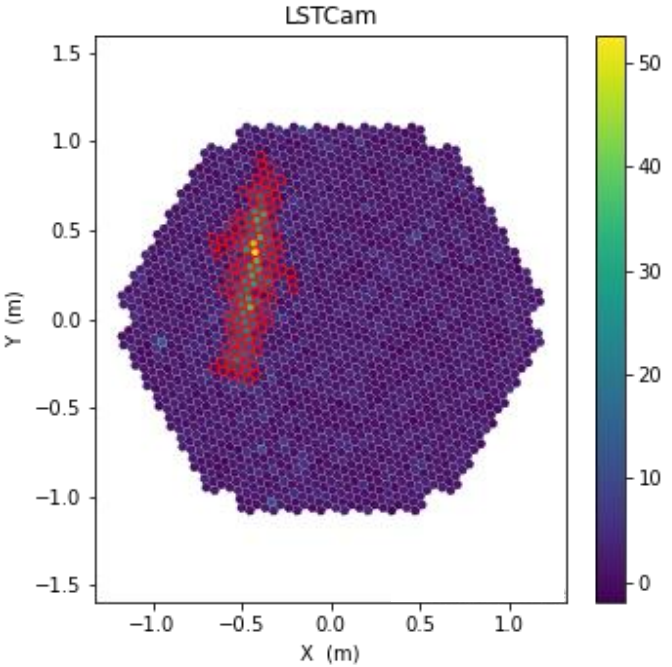
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



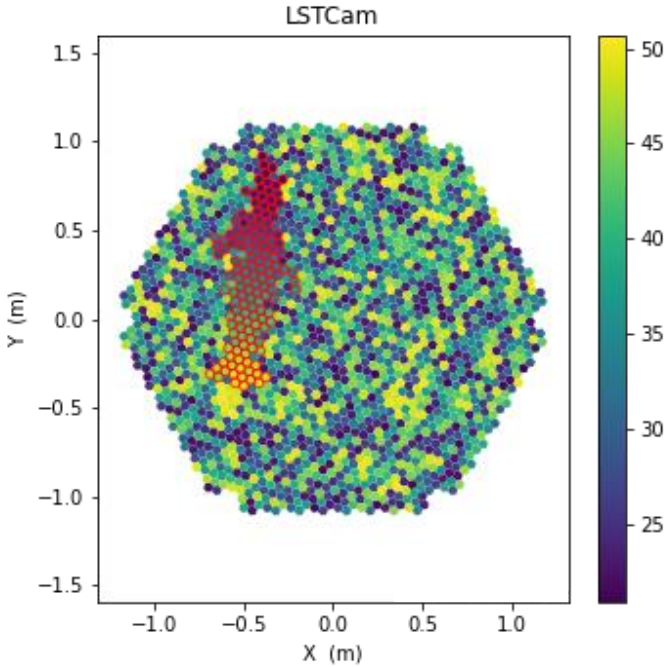
Time/ns



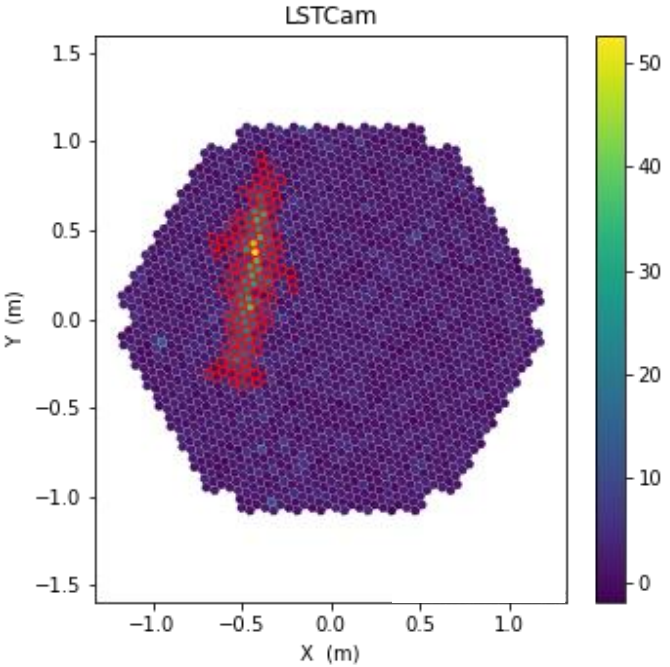
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



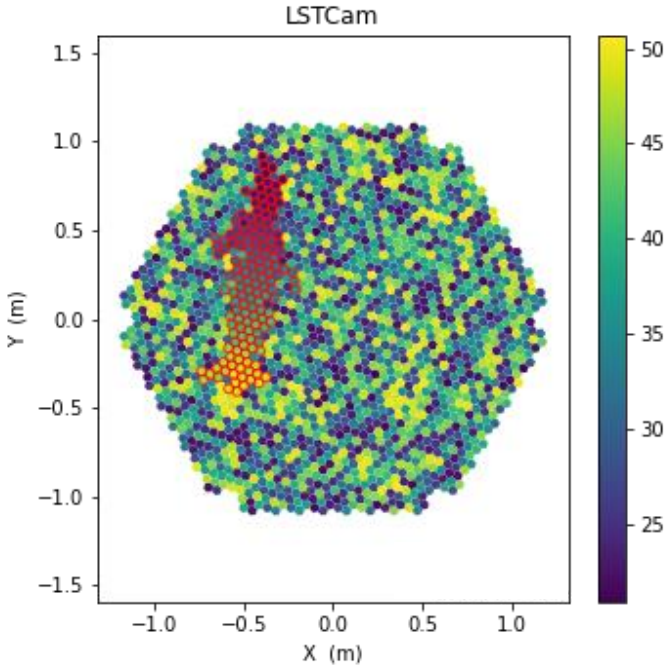
Time/ns



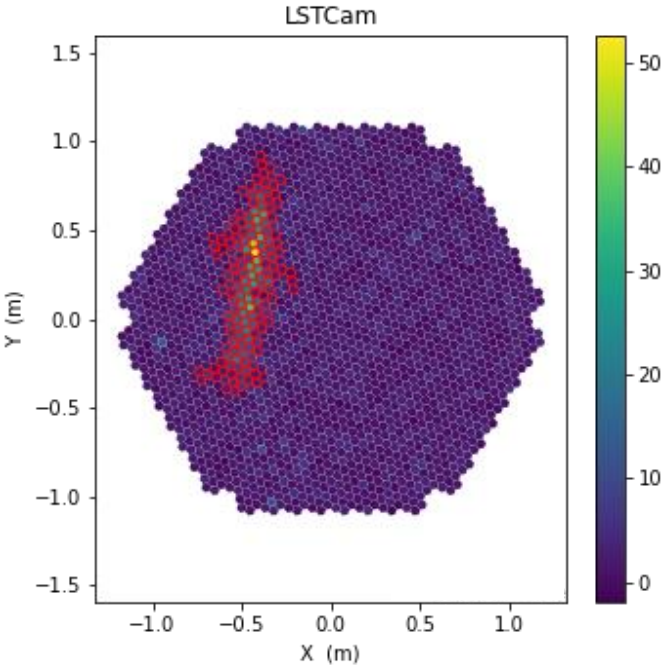
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



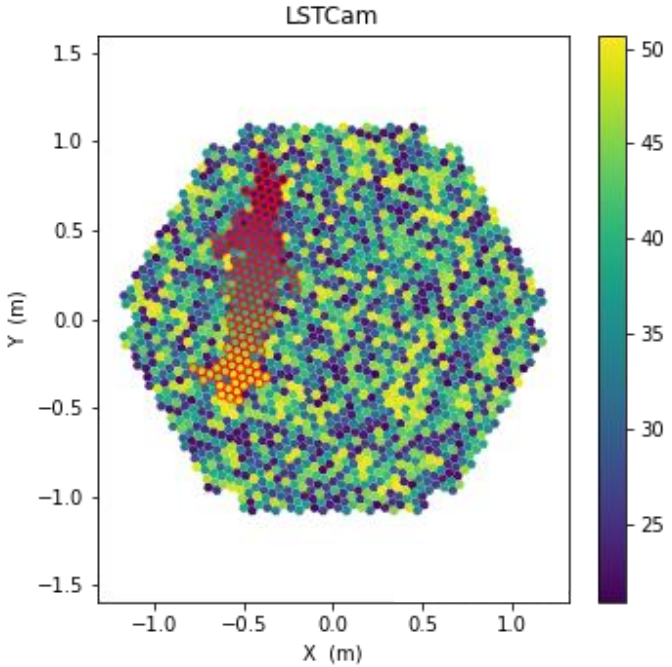
Time/ns



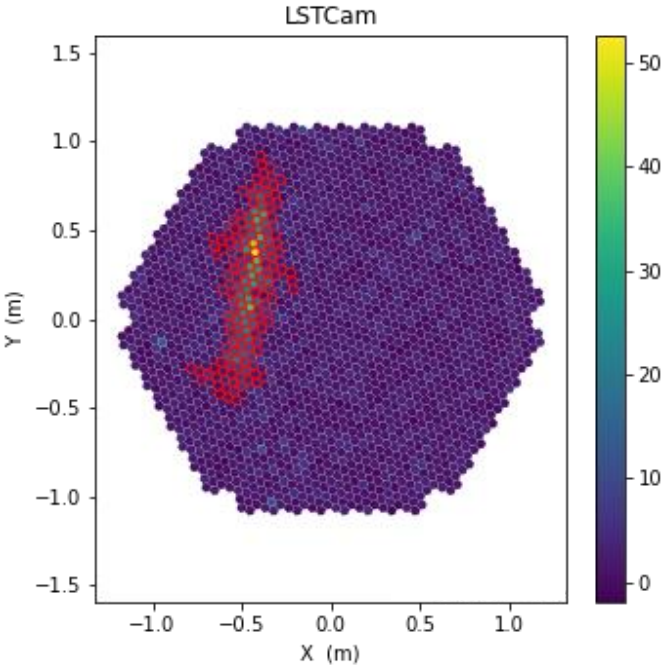
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



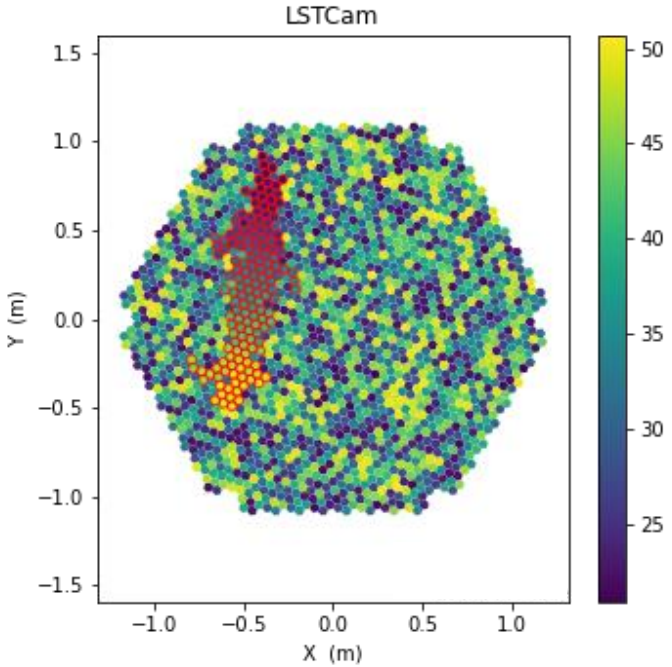
Time/ns



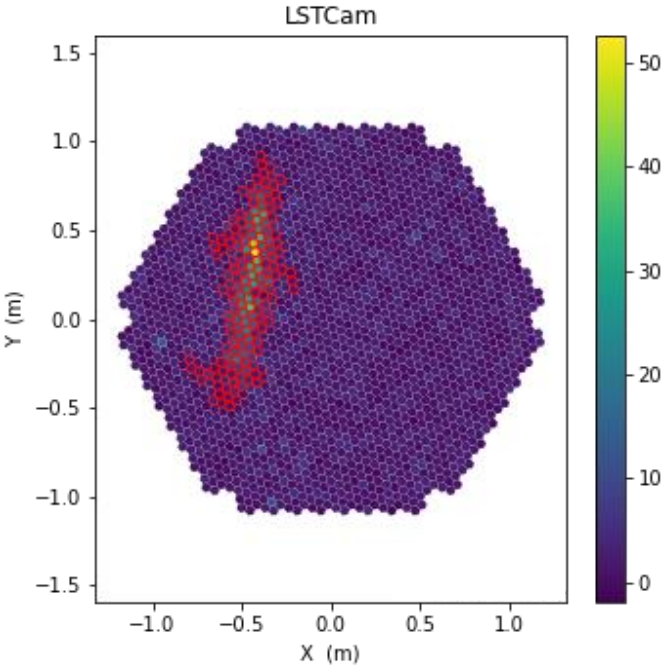
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



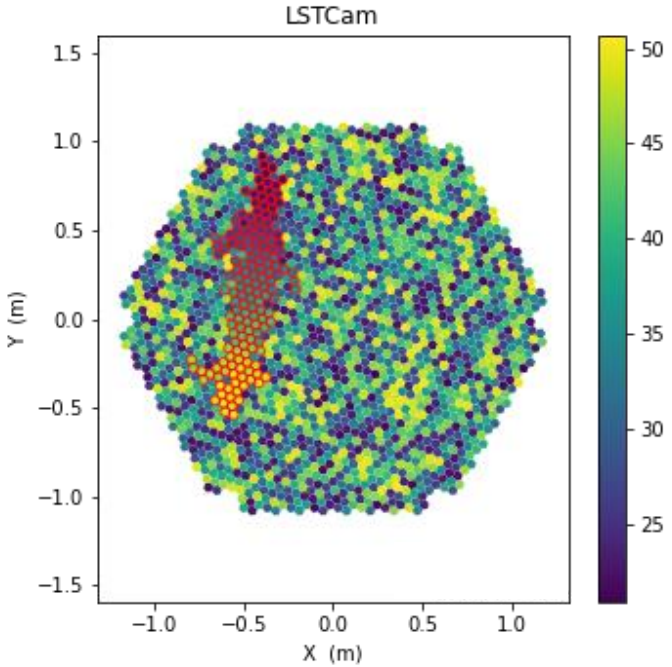
Time/ns



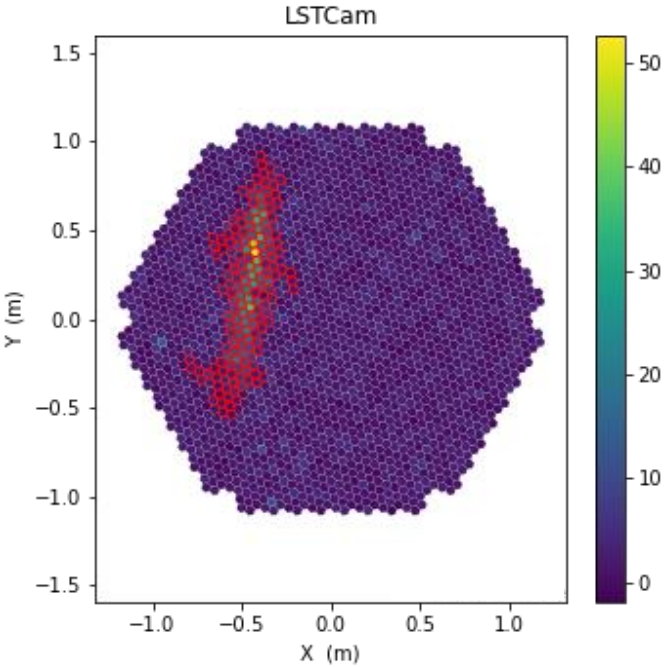
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold



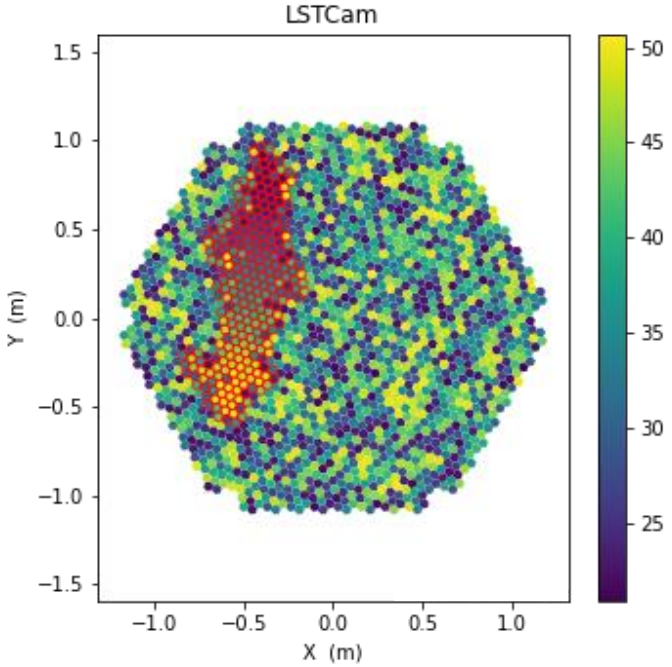
Time/ns



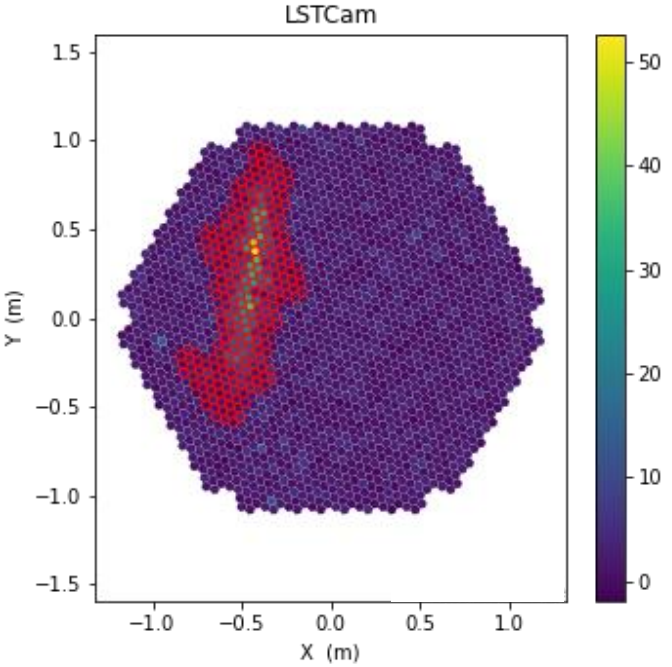
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold
3) At the end: Adding all neighbouring pixels



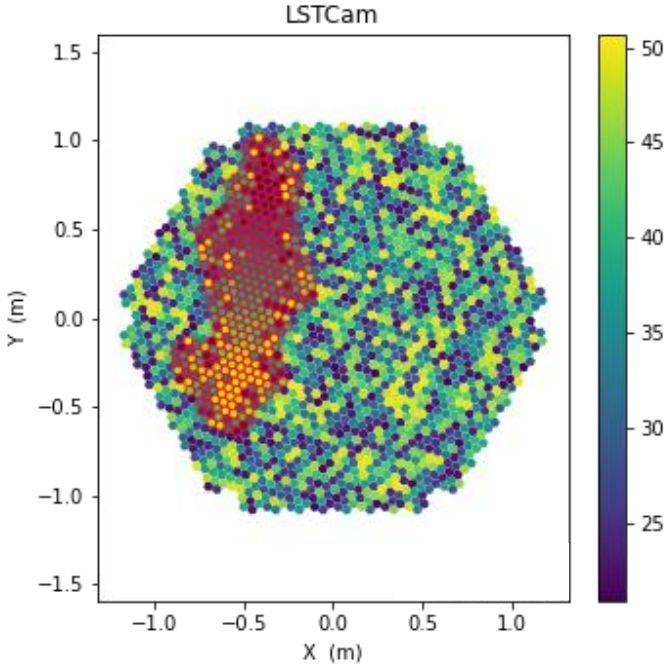
Time/ns



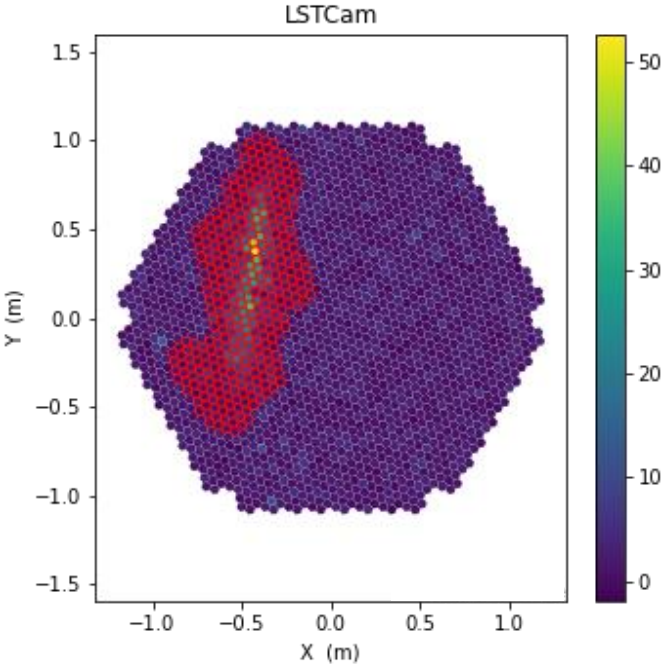
Photons

ArrivalTimesDataVolumeReducer

1) Normal Tailcuts cleaning on # Photons
2) Iteratively adding all neighbouring pixels with a time difference lower than a threshold
3) At the end: Adding all neighbouring pixels



Time/ns



Photons

TCDVR – ATDVR - Performance

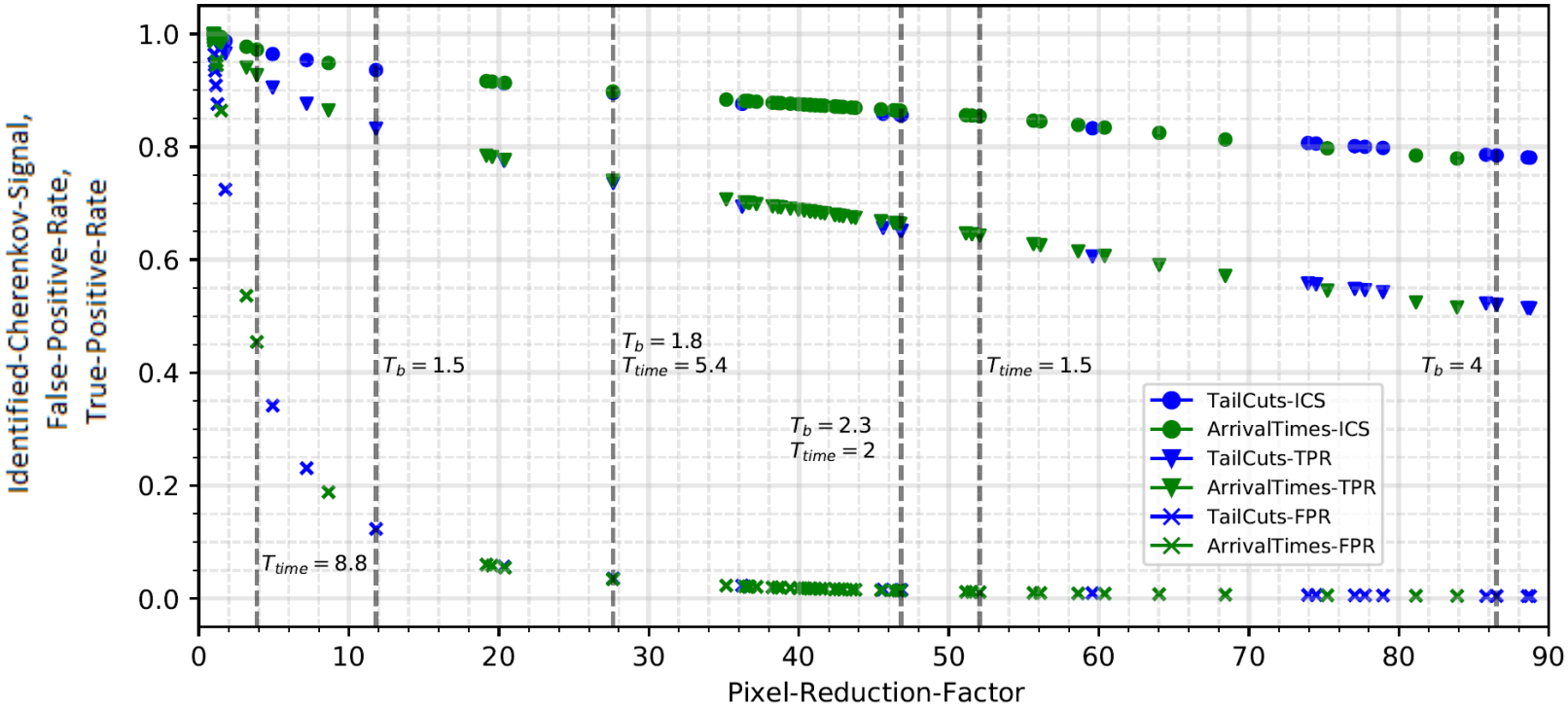
Quantitative Comparison on MC:

- First aim: Pixel-Reduction-Faktor of ~47 (~2.5 % of 1855 Pixels left)
- Some parameters were kept constant
- Set of values for T_b and T_{time} in a defined interval is drawn by means of a random selection

- Crab Nebula spectrum detected by HEGRA used for weights of gamma events
- Proton and Helium interpolated spectrum measured by DAMPE used for weights of proton events

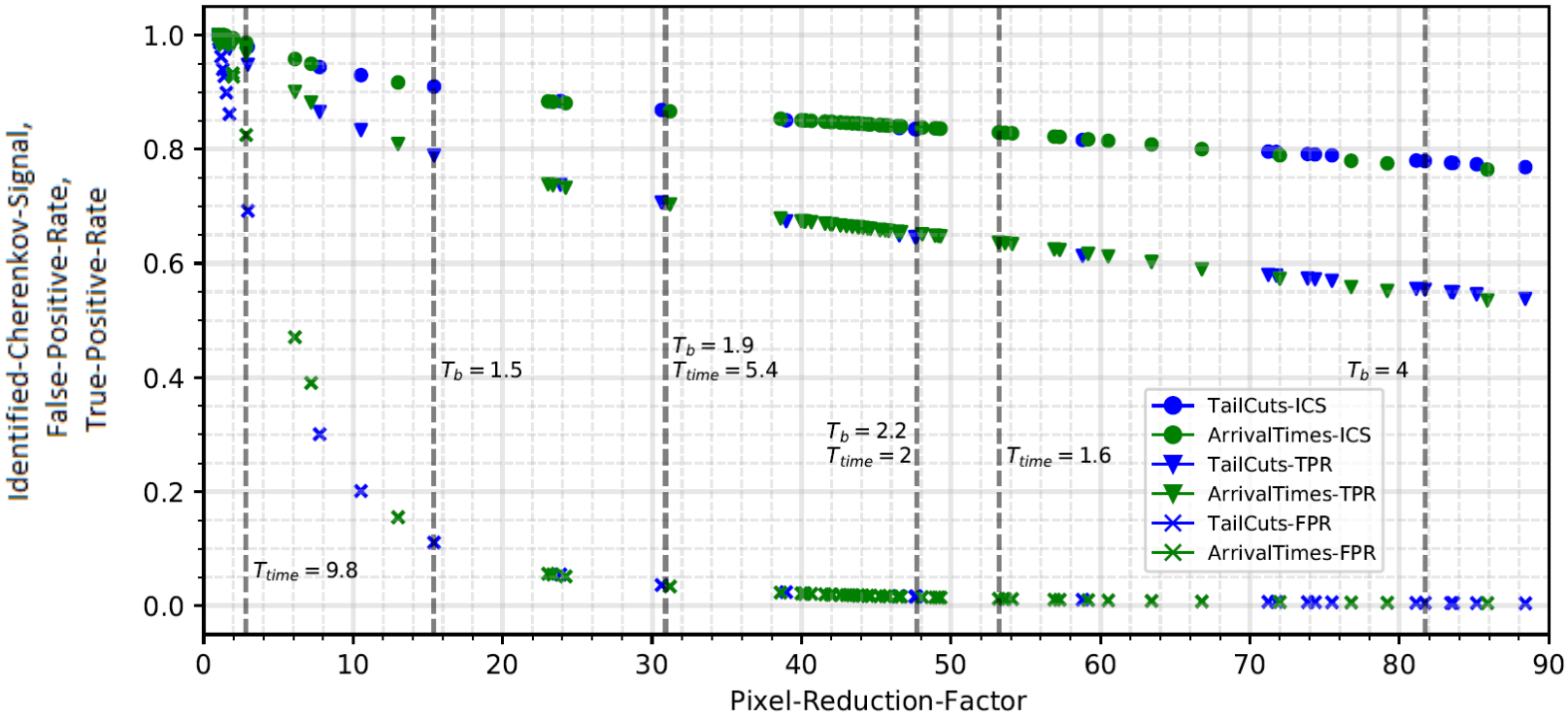
	TailCutsDVR	**ArrivalTimesDVR**
Core-Threshold T_c	8 p.e.	8 p.e.
Boundary-Threshold T_b	Variable	4 p.e.
Minimum core-neighbors	1	1
Keep isolated pixels	False	False
Dilate	1	1
Minimum time-neighbors	Not used	1
Time limit T_{time}	Not used	Variable

TCDVR – ATDVR - Performance



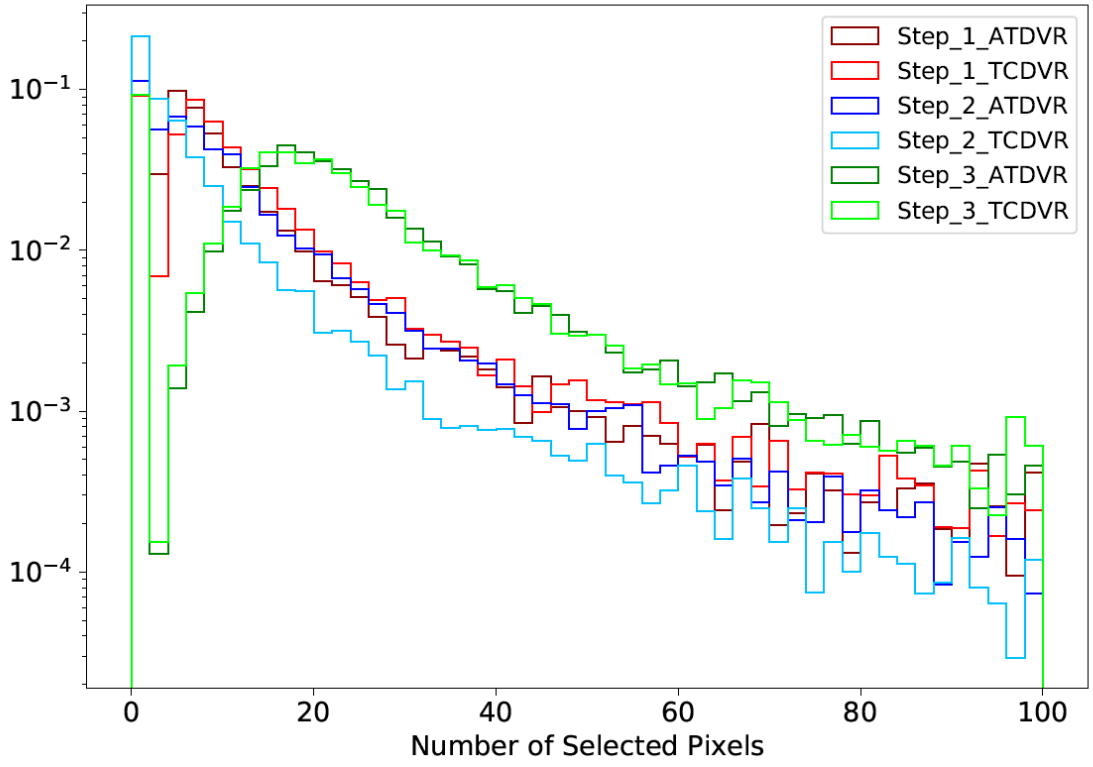
[20deg_Gamma_Prod5_MC]

TCDVR – ATDVR - Performance



[20deg_Proton_Prod5_MC]

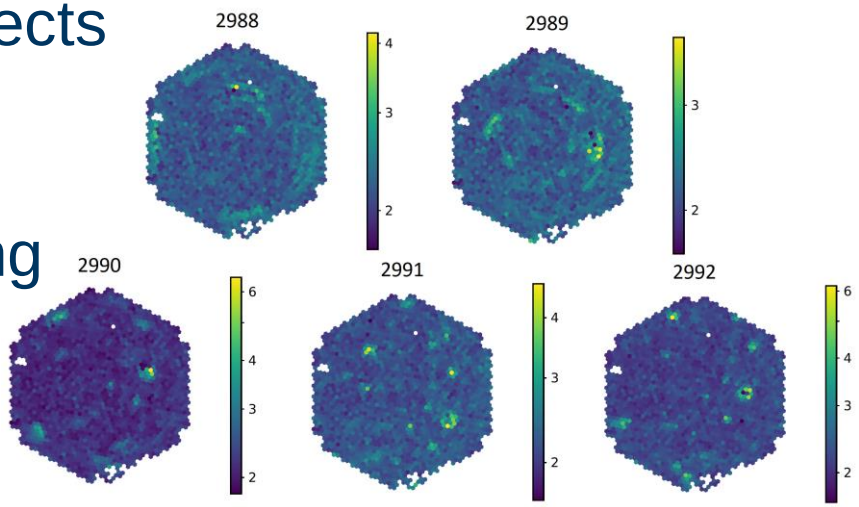
TCDVR – ATDVR - Performance



Crab Observation from 2020-11-21

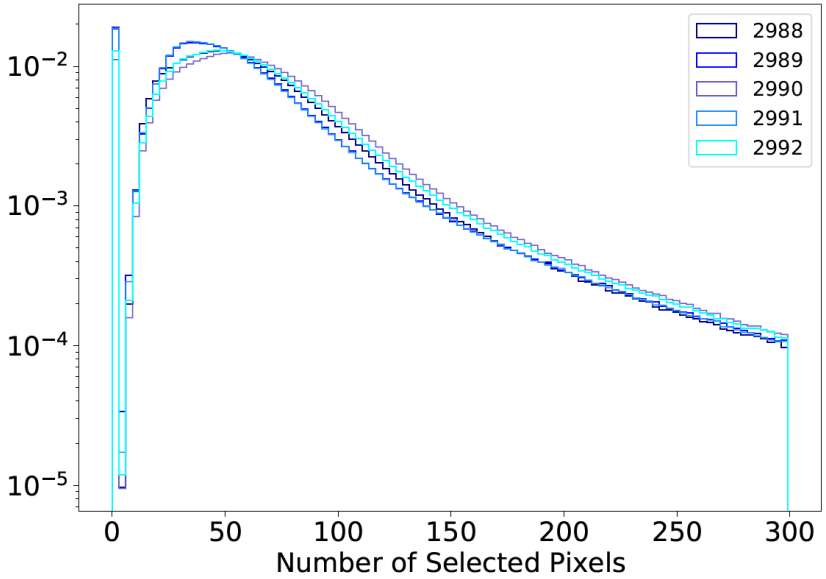
- Run 2988-2992 with ~1.6h Observation time
- 7-9 million events/file

- ➢ Flatfield Events: To reduce effects of image artifacts from the detector
- ➢ Interleaved Pedestals: Mapping the NSB



Crab Observation from 2020-11-21

Pixel selection with the ATDVR:

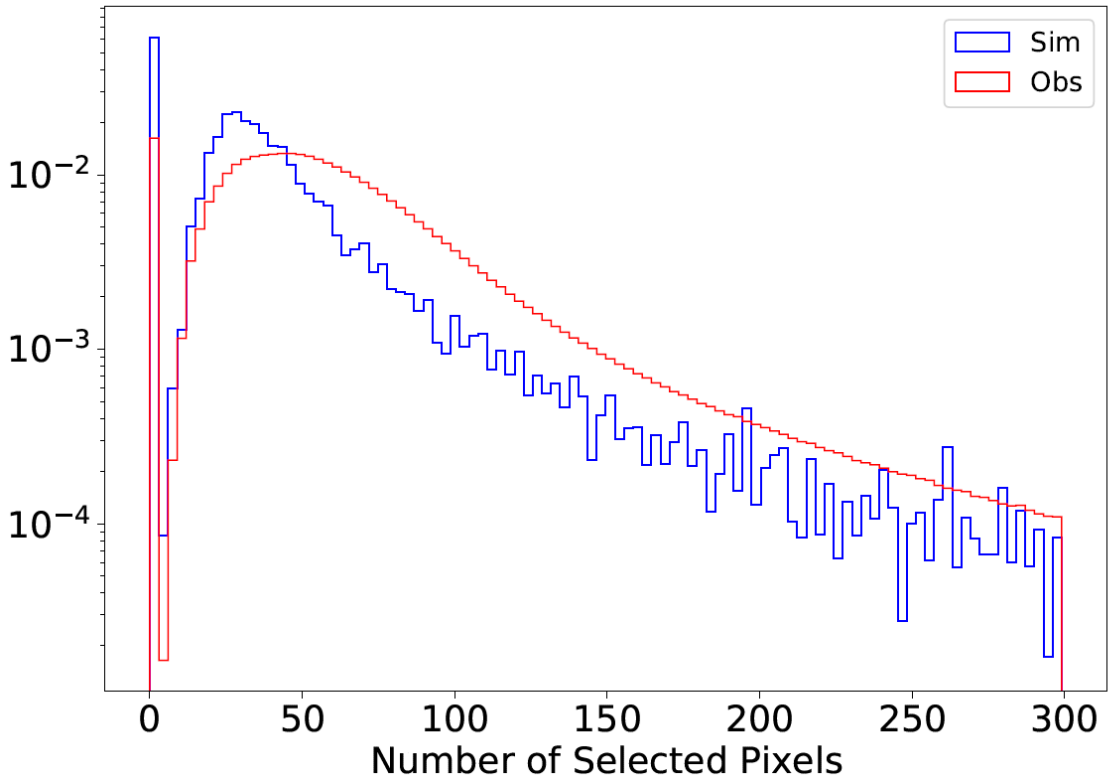


	ArrivalTimesDVR
Core-Threshold T_c	8 p.e.
Boundary-Threshold T_b	4 p.e.
Minimum core-neighbors	1
Keep isolated pixels	False
Dilate	1
Minimum time-neighbors	1
Time limit T_{time}	2.0

- Reduction per file from ~97.8 GByte to ~8.5 GByte
- PRF ~ 28

Crab Observation from 2020-11-21

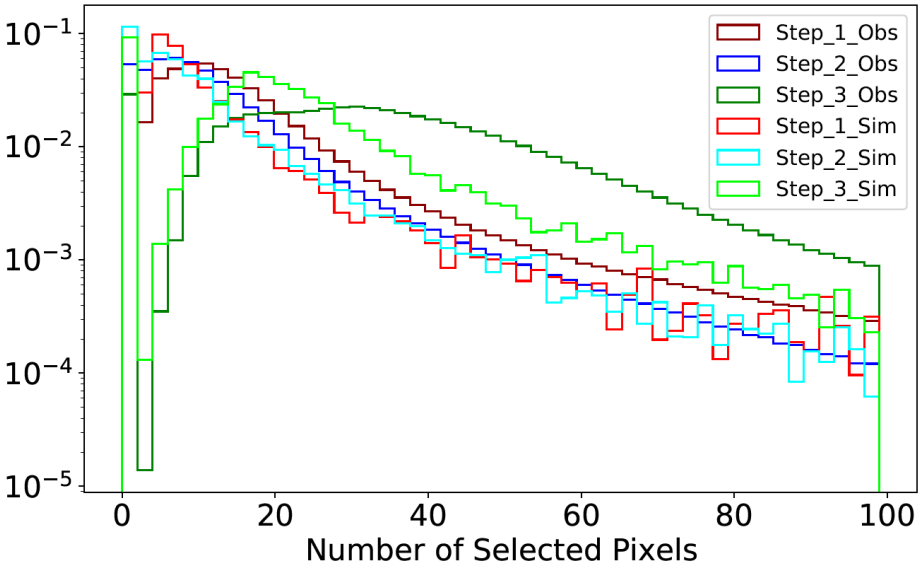
Pixel selection with the ATDVR:

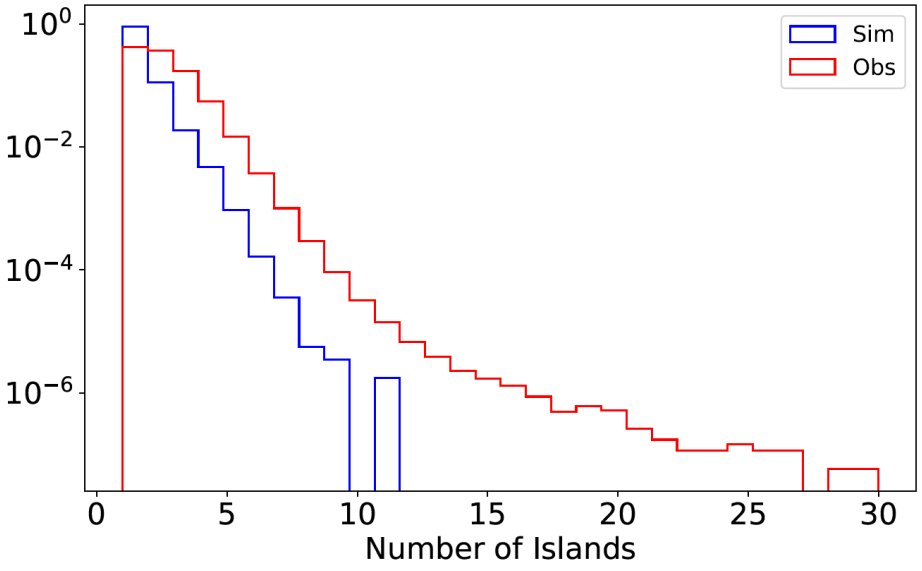


[weighted proton MC]

Crab Observation from 2020-11-21

Pixel selection with the ATDVR:





[weighted proton MC]

Crab Observation from 2020-11-21

Pixel selection with the ATDVR:

Approaching MC distribution of selected Pixels per image:

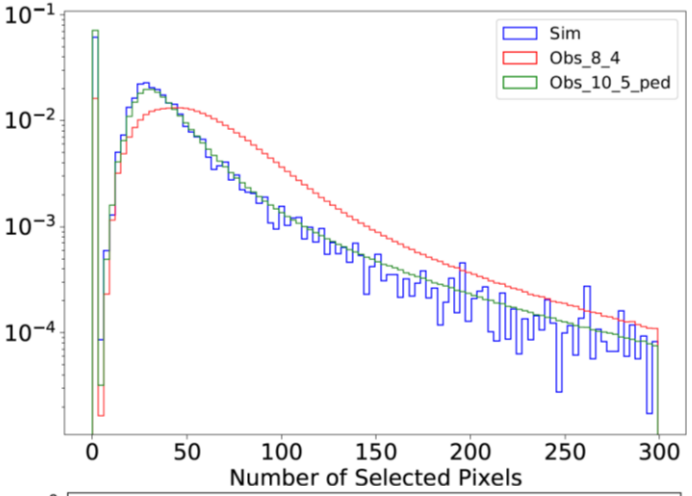
- Adapt $T_c = 10$ and $T_b = 5$
- Use interleaved pedestals for dynamic pixel thresholds

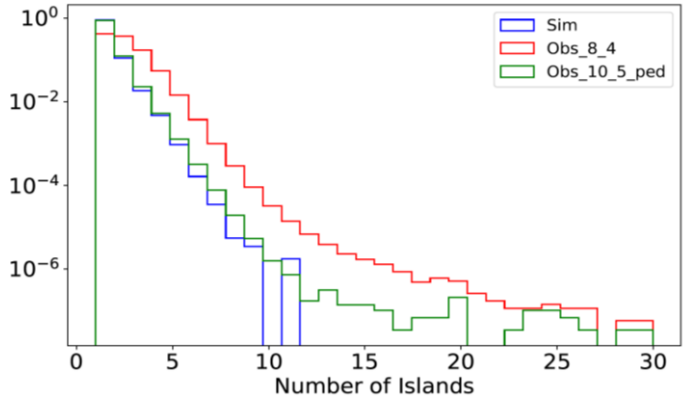
$$T_{ped} = \mu_{ped} + 2.5 * \sigma_{ped}$$

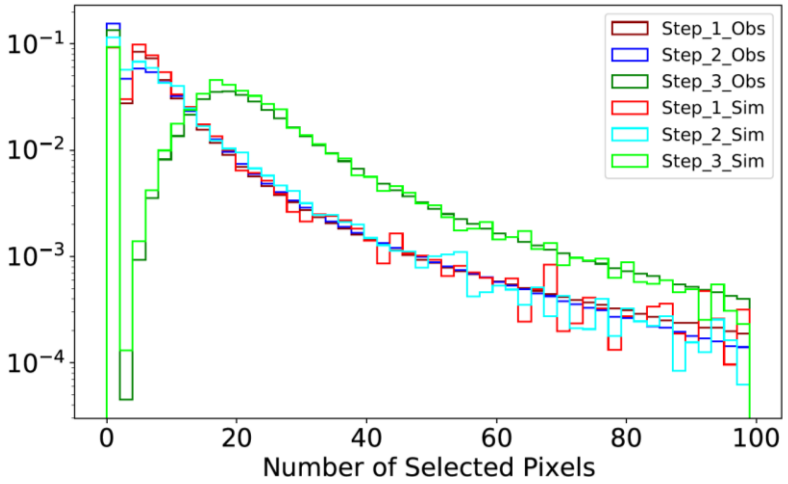
- Use T_{ped} if T_c is below T_{ped}

Crab Observation from 2020-11-21

Pixel selection with the ATDVR + adaptions:







- Reduction per file from ~8.5 GByte to ~6.8 GByte
- PRF ~ 42

[weighted proton MC]

Impact on Higher Level Analysis

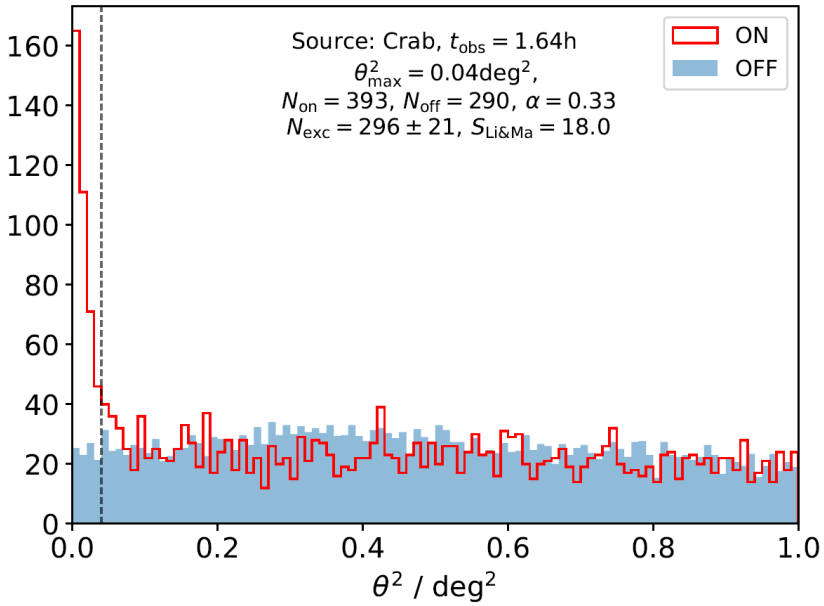
- Θ^2-Plot verification whether the analysis chain performs correctly
- Observations: Crab Nebula from 21.11.2020
- ArrivalTimesDVR
- Tailcuts Cleaning (8 | 4)
- DL2 processed with aict-tools (ctapipe processed MCs for RF training)
- Not optimized parameters

➢ Should have no difference by definition

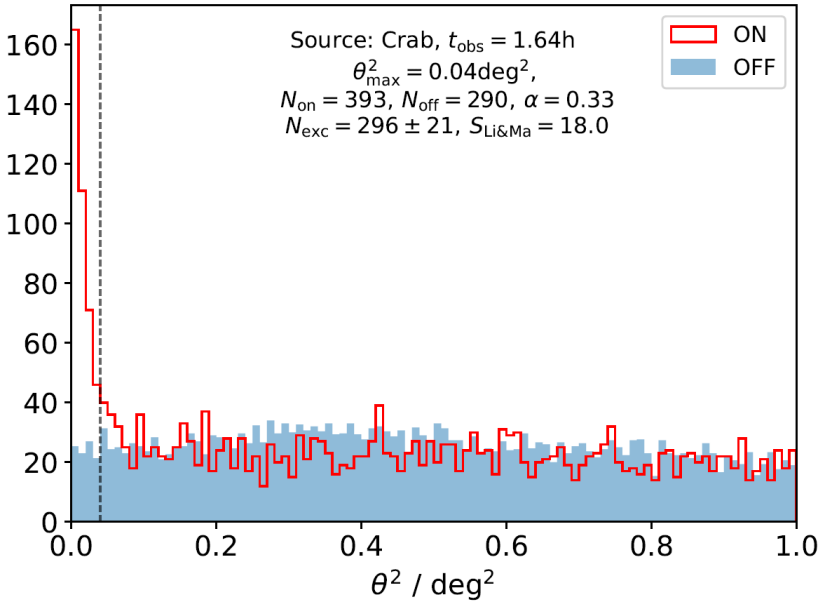
	ArrivalTimesDVR
Core-Threshold T_c	8 p.e.
Boundary-Threshold T_b	4 p.e.
Minimum core-neighbors	1
Keep isolated pixels	False
Dilate	1
Minimum time-neighbors	1
Time limit T_{time}	2.0

Impact on Higher Level Analysis

Without DVR



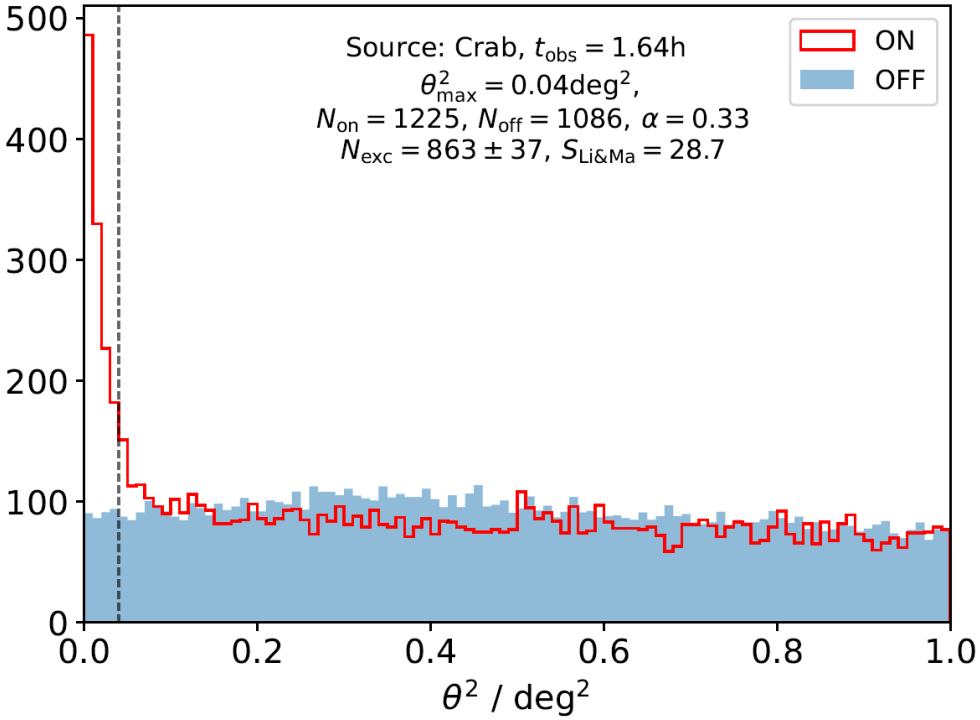
With DVR



[Crab Observation from 21.11.2020]

Impact on Higher Level Analysis

With Adaptions for DVR and cleaning:



[Crab Observation from 21.11.2020]

Summary

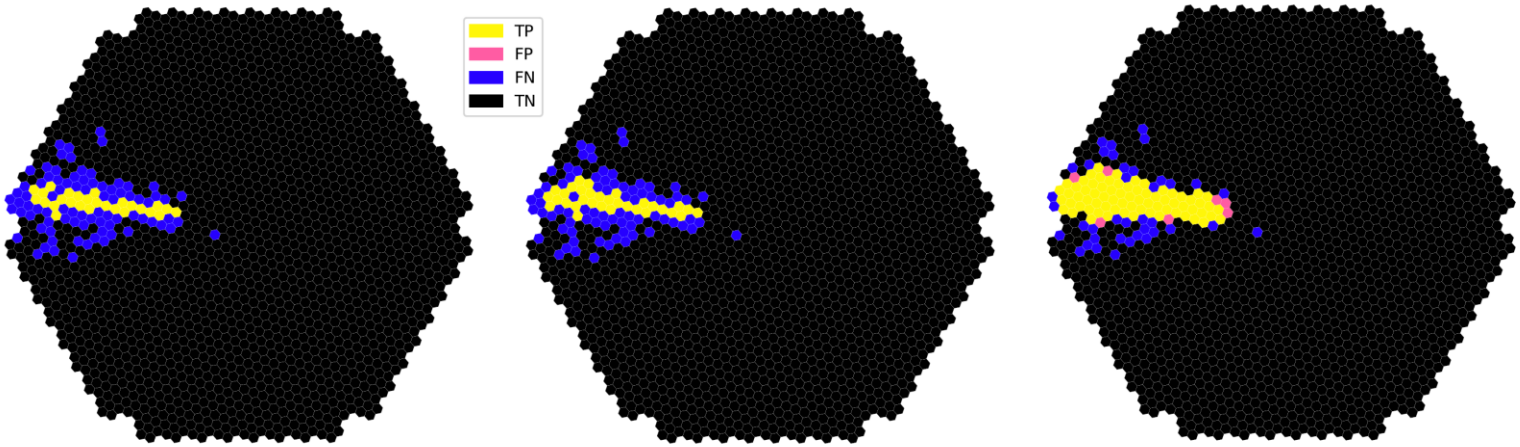
- TCDVR and ATDVR were tested and compared with MC data – perform equally well
- For PRF of 46: 0.83 ICS and 0.016 FPR on weighted simulated proton showers
- Employing ATDVR with same thresholds on Crab Nebula Observation: PRF of 28 and reduce data from 97.8 GByte to 8.5 GByte
- Increase PRF to 42 and decrease file size to 6.8 GByte with adapted thresholds
- Verifying used analysis chain by comparing Θ^2-Plots with and without DVR-step and detecting Crab Nebula with 18σ
- Detection increased to 28.7σ by using adapted cleaning thresholds

Outlook

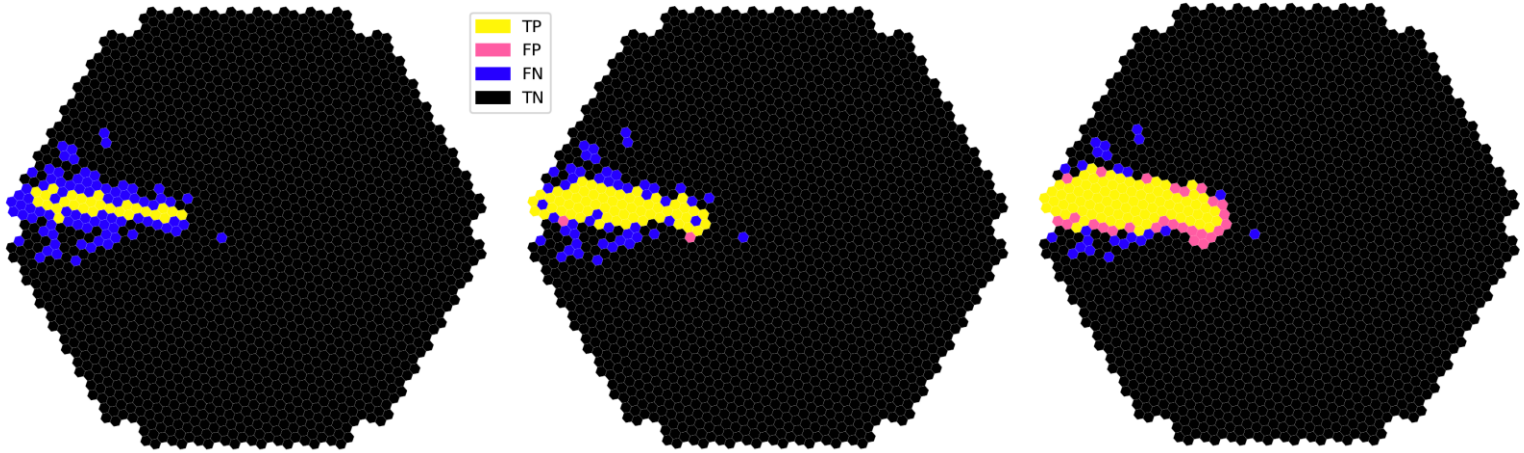
- Apply TCDVR to Crab Observations and compare it to the results obtained by the ATDVR
- Impact on higher data levels for different image cleaning algorithms
- Another way must be developed to include the interleaved pedestals events
- Vary other parameters
- Different benchmark criteria are needed (e.g. robustness, computation time, complexity etc.)
- General reference and training datasets for benchmarking must be defined

Backup TCDVR – ATDVR

TCDVR:



ATDVR:



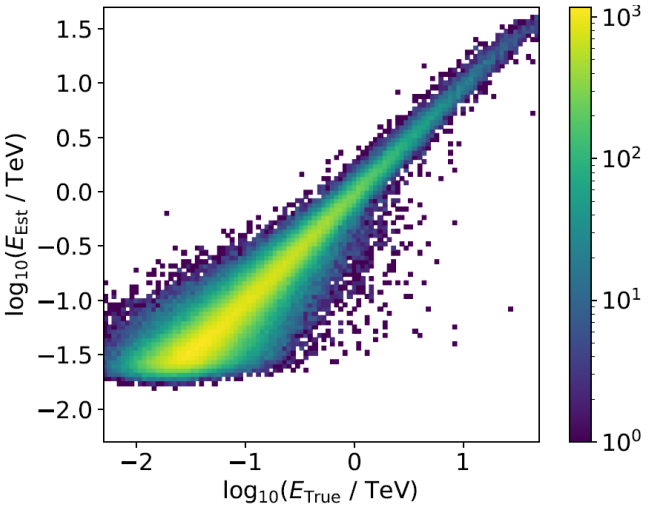
Backup – MC simulation properties

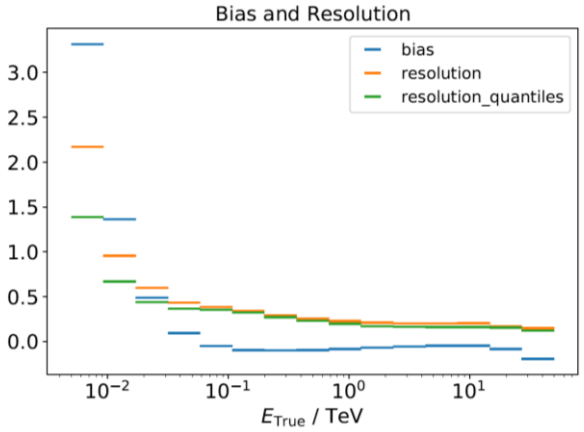
	Diffuse Gamma	Point-like Gamma	Proton	Elektron
Energy/TeV	0.005 - 50	0.005 - 50	0.001 - 100	0.005 - 5
Zenith angle/°	20	20	20	20
View cone/°	6	0	10	6
Spectral index	-2	-2	-2	-2

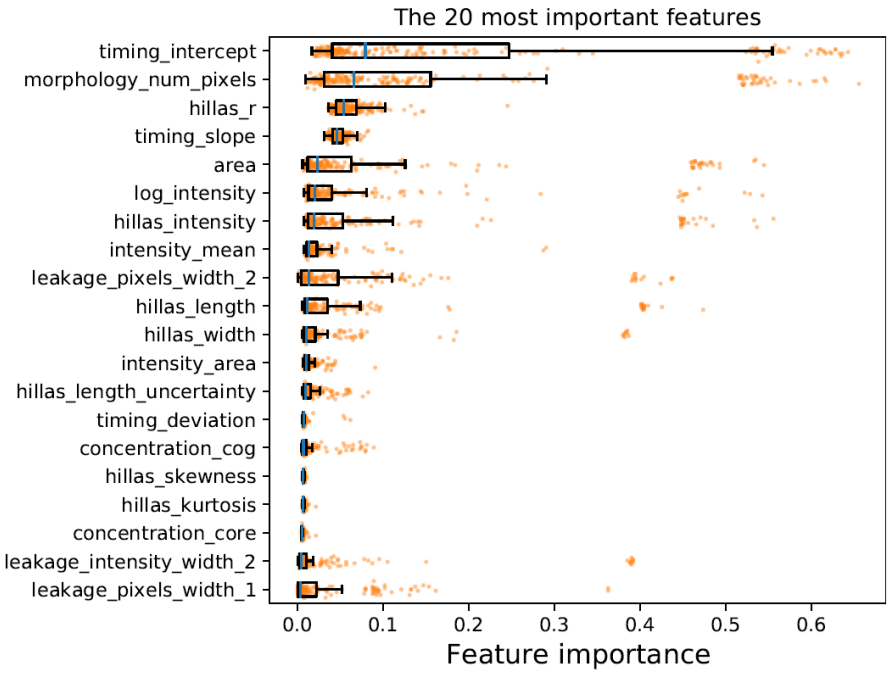
Backup – Quality Cuts

Parameter	Criteria
intensity	≥ 50
leakage_1_intensity	≤ 0.2
leakage_2_intensity	≤ 0.2

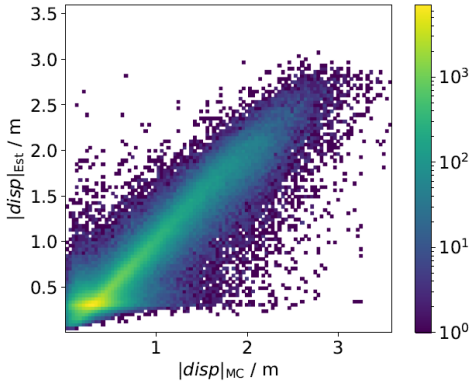
Backup – Energy Estimation

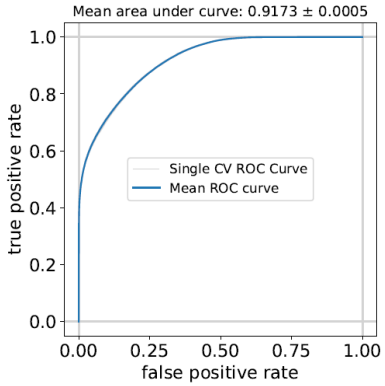


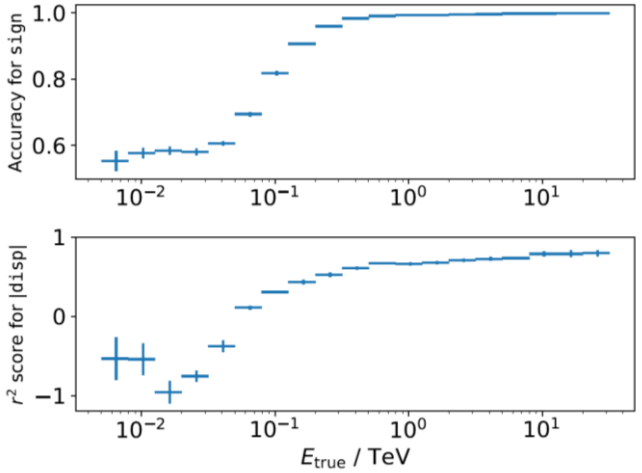


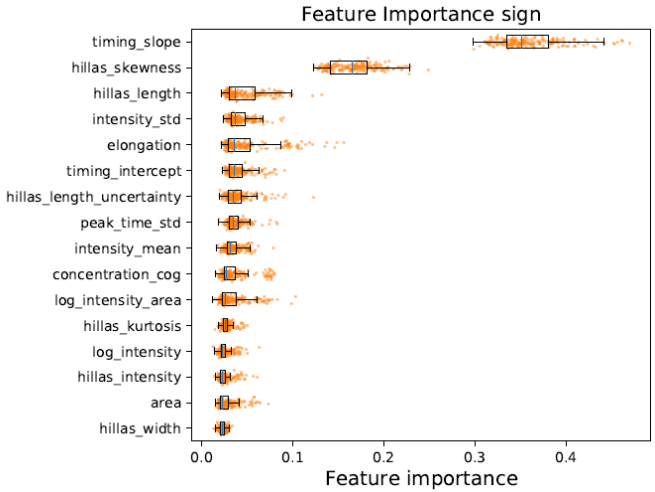


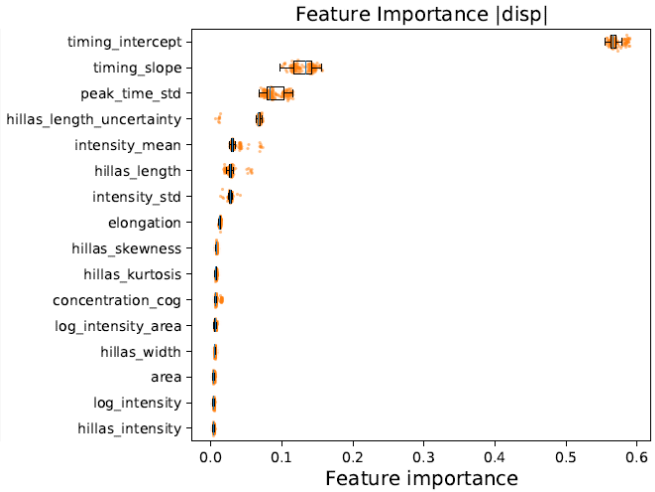
Backup – Origin Reconstruction











Backup – Background Separation