ASTRI Astrofisica con Specchi a Tecnologia Replicante Italiana

ASTRI data handling and archiving

S. Lombardi, INAF-OAR and ASI-SSDC, Rome, Italy

and L. A. Antonelli, C. Bigongiari, M. Cardillo, S. Gallozzi,

F. Lucarelli, M. Perri, F. G. Saturni, A. Stamerra, V. Testa

for the CTA ASTRI Project

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

Universidade de São Paulo

NORTH-WEST UNIVERSITY YUNIBESITI YA BOKONE-BOPHIRIMA NOORDWES-UNIVERSITEIT

agenzia spaziale italiana

SSDC SPACE SCIENCE DATA CENTER

Asterics

Outline

- ASTRI Project
- ASTRI Data Center
- ASTRI Data Handling activities:
  - ASTRI data reduction and analysis software
  - ASTRI MC simulations
  - ASTRI archive system
- Summary and outlook

ASTRI Project (in a nutshell)

- **Sub-project within Cherenkov Telescope Array (CTA)** led by INAF
- End-to-end prototype of the **CTA Small-Size Telescopes** (SSTs) with a dual-mirror optics design: the **ASTRI-Horn telescope**, installed at Mt. Etna (Italy) (verification phase in the last 2 years, scientific validation phase by fall 2019)
- **Mini-array of 9 ASTRI telescopes** to be deployed in Tenerife and proposed as a pathfinder of the CTA Observatory (joint efforts with Italian, Brazilian, and South African institutes, within CTA)
- Final aim: contribution to the installation of a considerable amount of (70) CTA SSTs

ASTRi-Horn telescope @ Serra La Nave (Mt. Etna, Sicily)

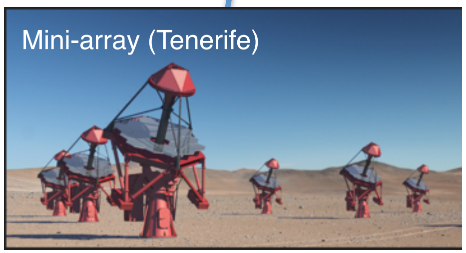
Artistic view of the ASTRI mini-array in Tenerife

ASTRI Data Handling overview

Prototype (Mt. Etna)

ASTRI is an INAF end-to-end project aimed at the realization of a prototype of a dual-mirror Cherenkov Telescope and of a mini array of 9 of such a telescopes.

The Project is inclusive of a complete *Data Handling System* and foresees an ASTRI Data Center.



Data Acquisition

On-site Archive

RAW, CAL, TECH, LUT, EVT, SCI

LEVEL A ANALYSIS (On-line/On-site Analysis)

LEVEL B ANALYSIS (Off-line/On-site Analysis)

OBSERVATORY

DATA CENTER (Rome)

ASTRI GATEWAY

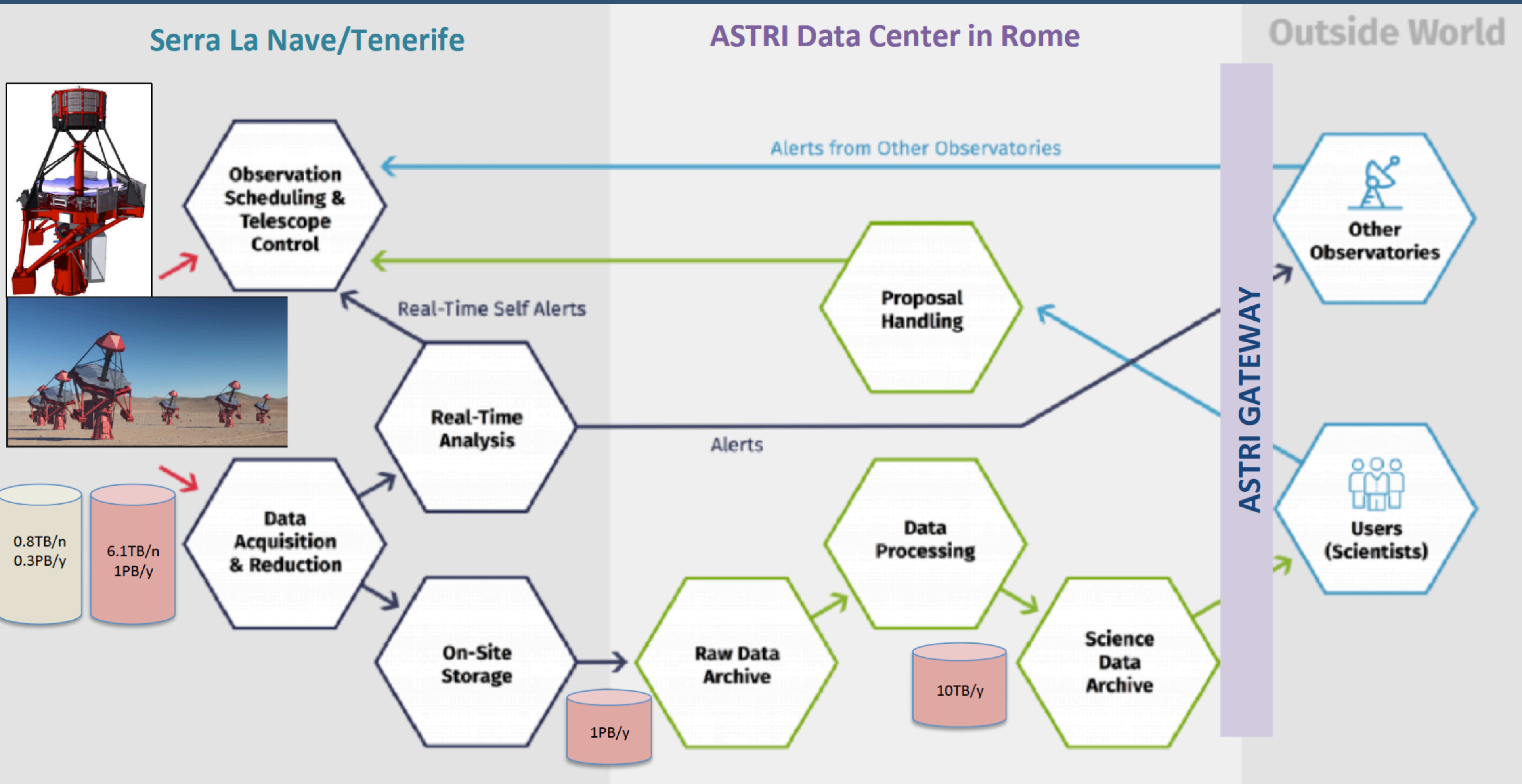
Off-Site Archive

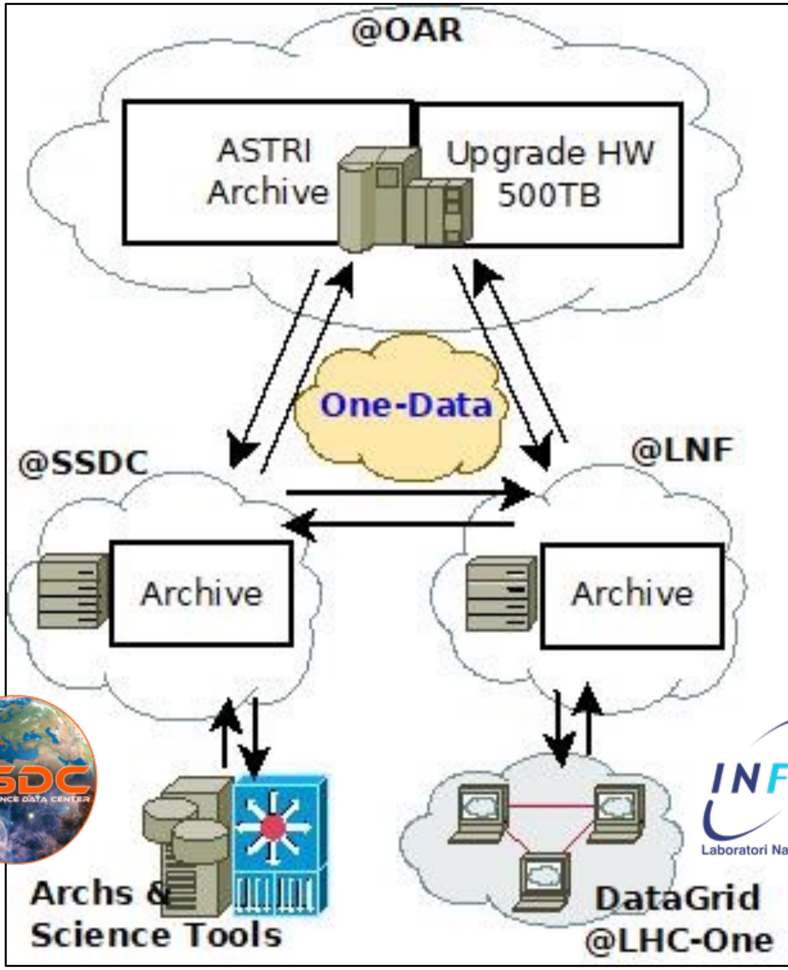
RAW, CAL, TECH, MC, EVT, SCI, VO

LEVEL C ANALYSIS (Off-line/Off-site Analysis)

§ 4

ASTRI data flow





ASTRI Data Center in Rome

- **Pipelines** developed in INAF and SSDC also within ***H2020 Project ASTERICS*** in a end-to-end approach.
- **Archive concept:** developed within the ***H2020 Project INDIGO-DataCloud*** as a distributed archive.
- Data already present in the Archive: ASTRI MC data, real data from ASTRI prototype, scientific simulations from ACDC.
- **3 sites involved: @INAF-OAR** (as main archive of ASTRI prototype)**, @INFN-LNF** to access the GRID**, @SSDC-ASI** to provide final user with a data access in a scientific & MWL environment.
- **Science Gateway:** to provide access to users from the preparation and submission of Observing Proposals to final scientific data.

ASTRI CTA Data Challenge Project

ACDC (ASTRI/CTA Data Challenge) is an INAF project (PI. P. Caraveo) carried out in the framework of the CTA and ASTRI programs to foster the Italian community of the TeV astronomy, developing know-how and experience for data-analysis in the light of the early science of the CTAO.
ACDC is includes 85 scientist (36 staff FTE + 9,6 non-staff FTE) in 9 INAF structures. The project officially started on September 2017 and will end in 2019.

Aims

- **End-to-end simulation of a realistic 3 years of observations of a sample of targets**:
  - 9 ASTRI SSTs in a realistic layout
  - E range = ~1 TeV – ~200 TeV
  - FoV ~ 10 deg^2
- Call for Proposal process
- Specific Pointing plan

Simulation and Analysis

- Simulation performed with *CORSIKA/sim_telarray*
- IRF generation with *A-SciSoft*
- Simulation of event-lists with *ctools* (v.1.5.2)
- Systematic analysis of all the observations to determine significance, mean flux and spectrum of all the simulated sources (with *ctools*)
- No blind source-detection / No temporal selection

Selected Targets

- **LS 5039**
- **LMC P3**
- **Sculptor**
- **Reticulum II**
- **Tucana II**
- **HESS J1748-248**
- **HESS J1018-589**
- **HESS J1825-137**
- **HESS J1303-631**
- **Vela X**
- **HESS J1632-478 and HESS J1634-472**
- **HESS J1833-105**
- **SNR G0.9+0.1**
- **MSH 15-52**
- **NGC 1068**
- **W28**
- **Westerlund 2**
- **Crab**
- **PKS 2155-304**

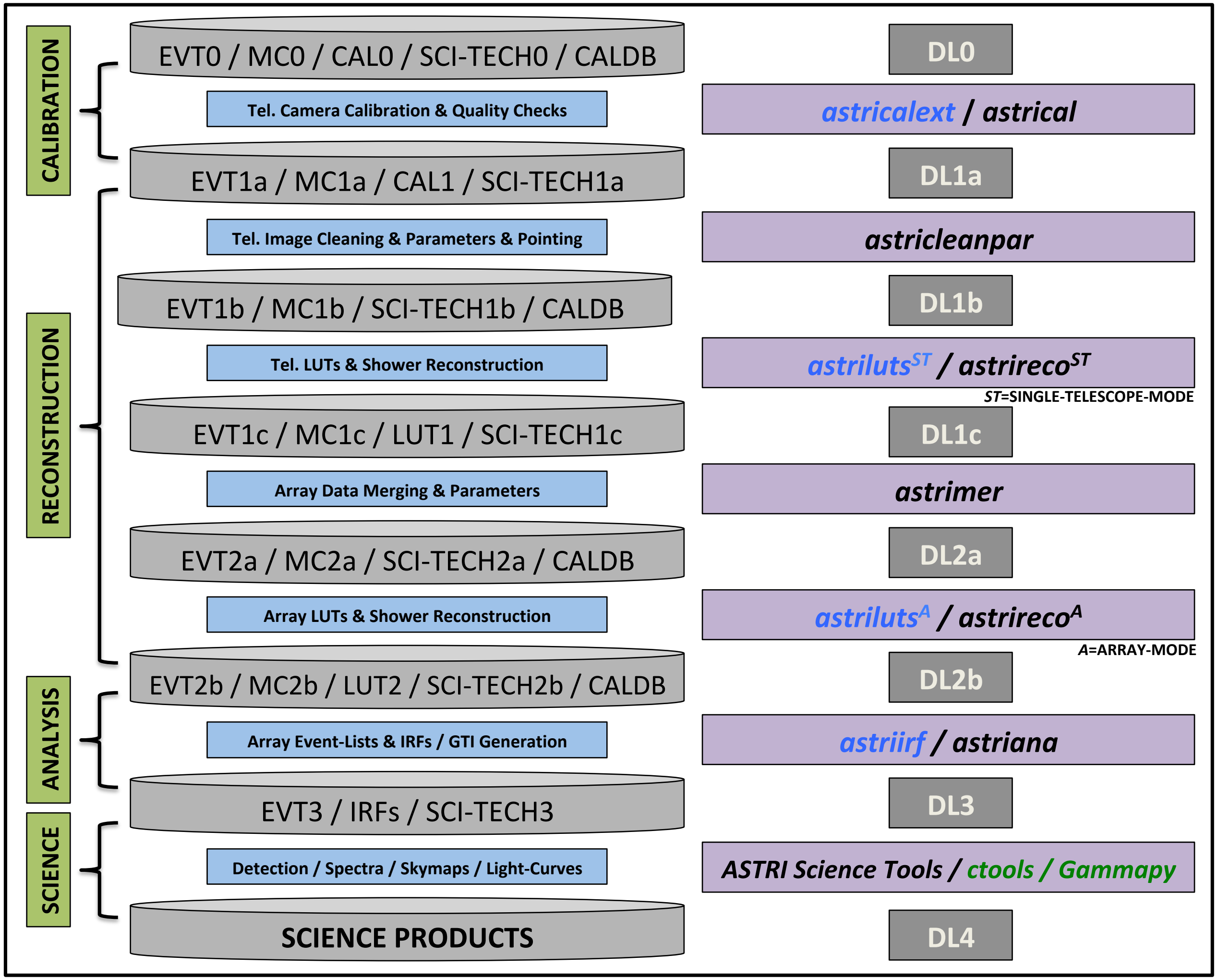
ASTRI Data Handling main activities:

- **On-site/off-site Archives and Pipelines**
- **MC simulations** (for performance assessment and real data reduction)

- Prototype data reduction (for commissioning/validation phases)
- Mini-array IRFs production (for science prospects and INAF ACDC Project)
- Utilization/testing of CTA Science tools (*ctools* and *Gammapy*)
- Real Time Analysis for ASTRI prototype
- Machine/Deep Learning activities
- Scientific Gateway for the ASTRI mini-array

- **Currently focus on ASTRI SST-2M prototype activities**
- **Soon focus on preparatory phase for the mini-array DH system**

ASTRI data reduction and analysis software



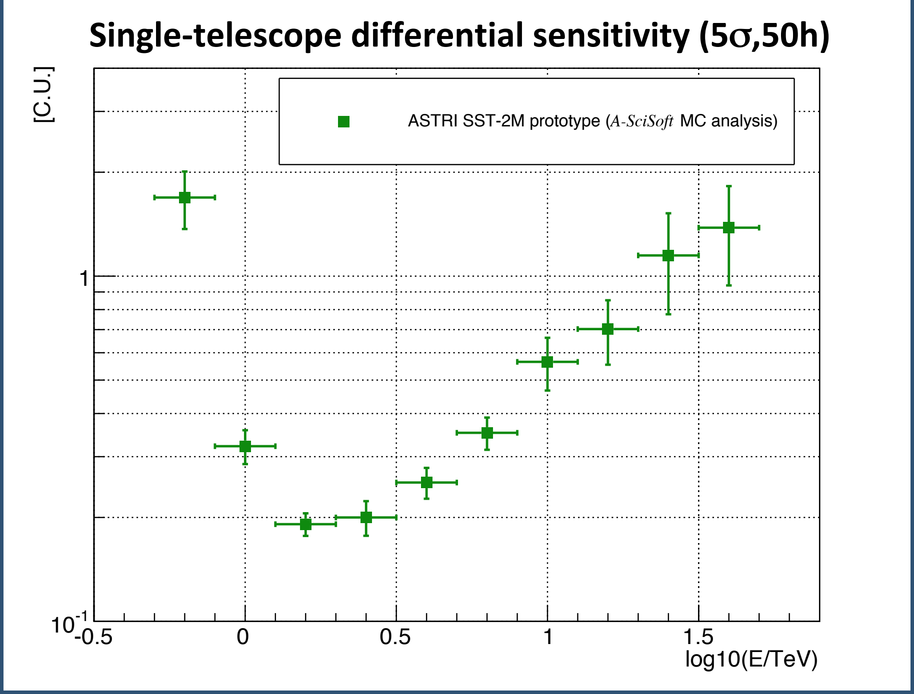
Breakdown stages; Basic functionalities; Auxiliary modules / **Pipeline modules**; I/O Data level.

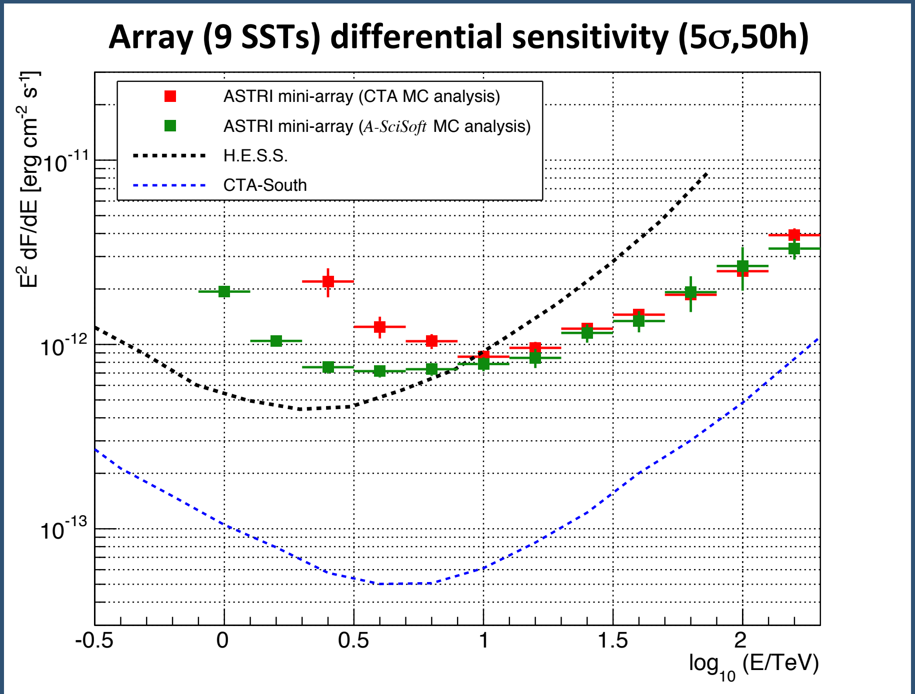
SPIE 991315 (2016); SPIE 107070 (2018);
Astronomy & Computing (in prep.)

A-SciSoft (astripipe):

- Real and MC ASTRI data reduction
- Single-telescope-wise and Array-wise data reduction
- On-line/on-site/off-site data reduction
- run on x86 / ARM CPUs & NVIDIA GPUs
- FITS data format from DL0 to DL4
- *C++/Python/CUDA*
- Independent (*auxiliary* and *pipeline*) modules
- *Python* pipeline wrapper
- CTA Science Tools compliance

- Validation tests from dedicated ASTRI MC simulations:
  - **End-to-end single-telescope & array MC data reduction validation**
    - Single-telescope sensitivity in line with (analytical) expectations
    - Array sensitivity in line with previous estimates*
      (achieved with an independent MC data analysis chain)

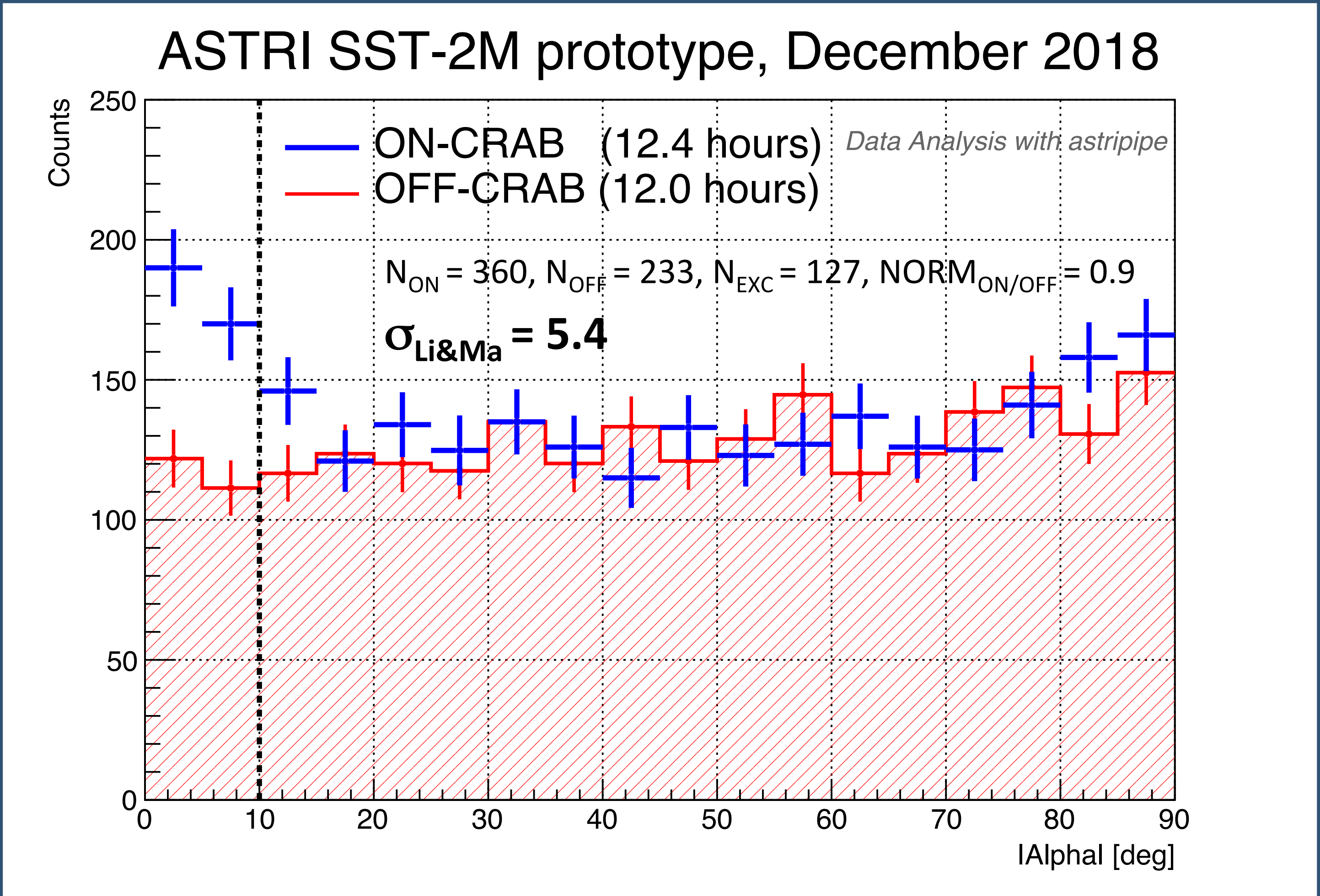




* *Di Pierro et al., J. Phys.: Conf. Ser. 718, 052008 (2016).*

Real prototype data reduction and analysis

- **First ever detection of an astrophysical source at TeV energies with a Cherenkov telescope in dual-mirror Schwarzschild-Couder configuration**
- **First detection of an astrophysical source by a CTA prototype telescope**



MC simulations software

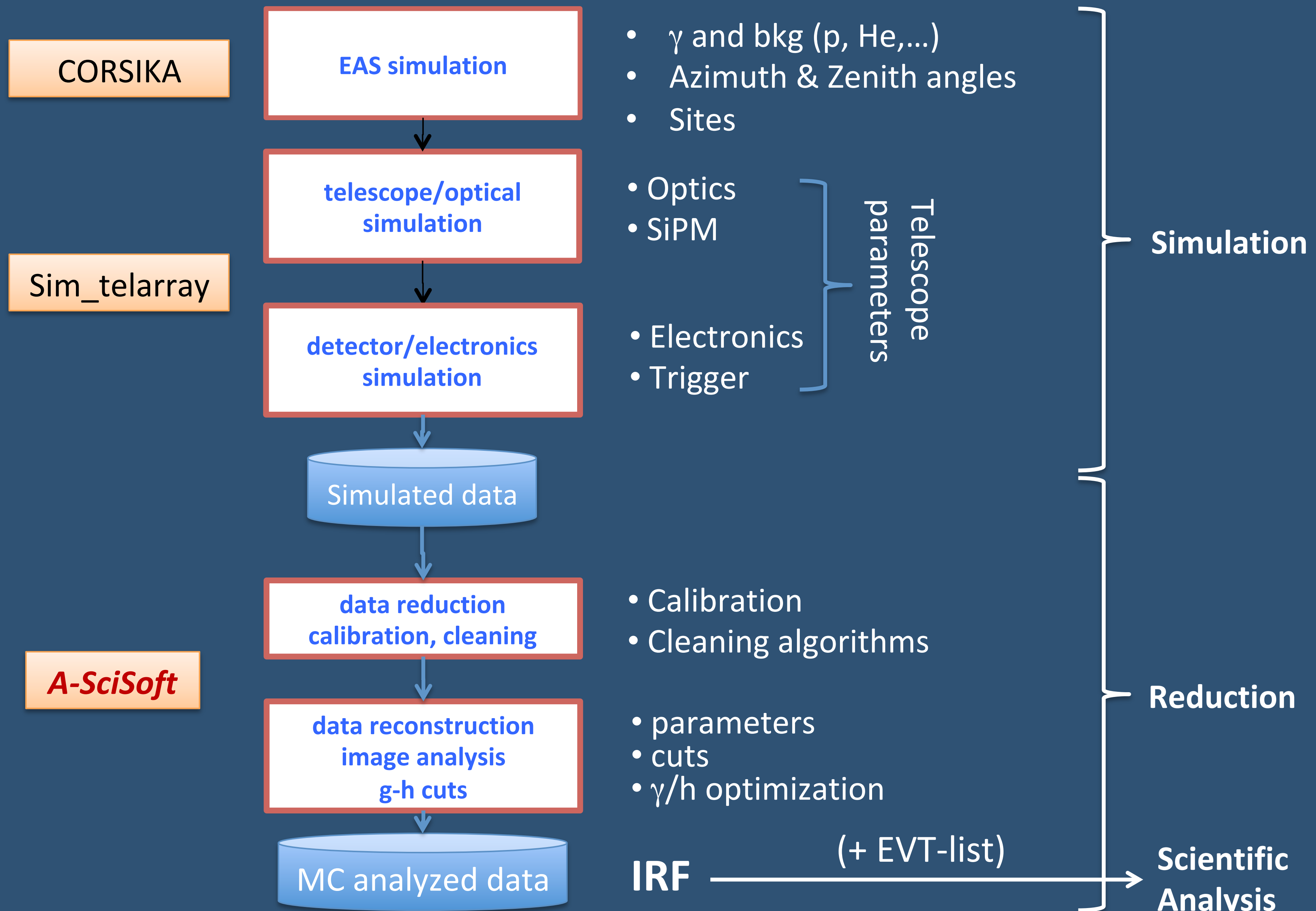
- ***CORSIKA*** (COsmic Ray Simulation for Kascade) is a program for detailed simulation of extensive air showers initiated by high energy cosmic ray particles by D.Heck and T.Pierog
  - D. Heck et al., Report FZKA 6019 (1998)
  - https://www.ikp.kit.edu/corsika/index.php
  - Used by many experiments in gamma-ray astronomy, neutrino astronomy and cosmic-ray physics
  - CORSIKA version 6.99 (presently in use in CTA)

- ***Sim_telarray*** is a program for detailed simulation of IACT by K. Bernlöhr
  - K. Bernlöhr, Astropart.Phys.30:149-158 (2008)
  - https://www.mpi-hd.mpg.de/hfm/~bernlohr/sim_telarray/
  - Extensively cross-checked against data from HEGRA and HESS arrays
  - Never used before CTA on dual mirror telescopes nor on telescopes not equipped with FADCs

- ***ASTRIconverter*** (part of the A-SciSoft software package)

- ***A-SciSoft*** (already discussed in previous slides)

MC simulations and analysis workflow



Computational and storage needs

Let's consider a realistic simulation of an array composed of 15 ASTRI-like telescopes to estimate its IRFs. We need to simulate at least gamma and proton events over the energy range of interest.

	E_{Min} [GeV]	E_{Max} [TeV]	E_{Slope}	R_{Max} [m]	ϑ_{Max} [deg]	Random points	<E> [TeV]
Gamma	100	330	1.5	1800	10.0	20	5.74
Proton	100	600	1.5	2400	10.0	20	7.75

We need at least 2×10^8 gamma and 2×10^9 proton showers to estimate the IRFs of this array

The simulation of these showers with *CORSIKA* 6.99 requires
~2×10^5 HS06 years with **~100 TB** of output data (number of files ~2×10^4)

The simulation of the telescope response with *sim_telarray* requires
~2×10^4 HS06 years with **~3 TB** of output data (number of files ~3×10^5)

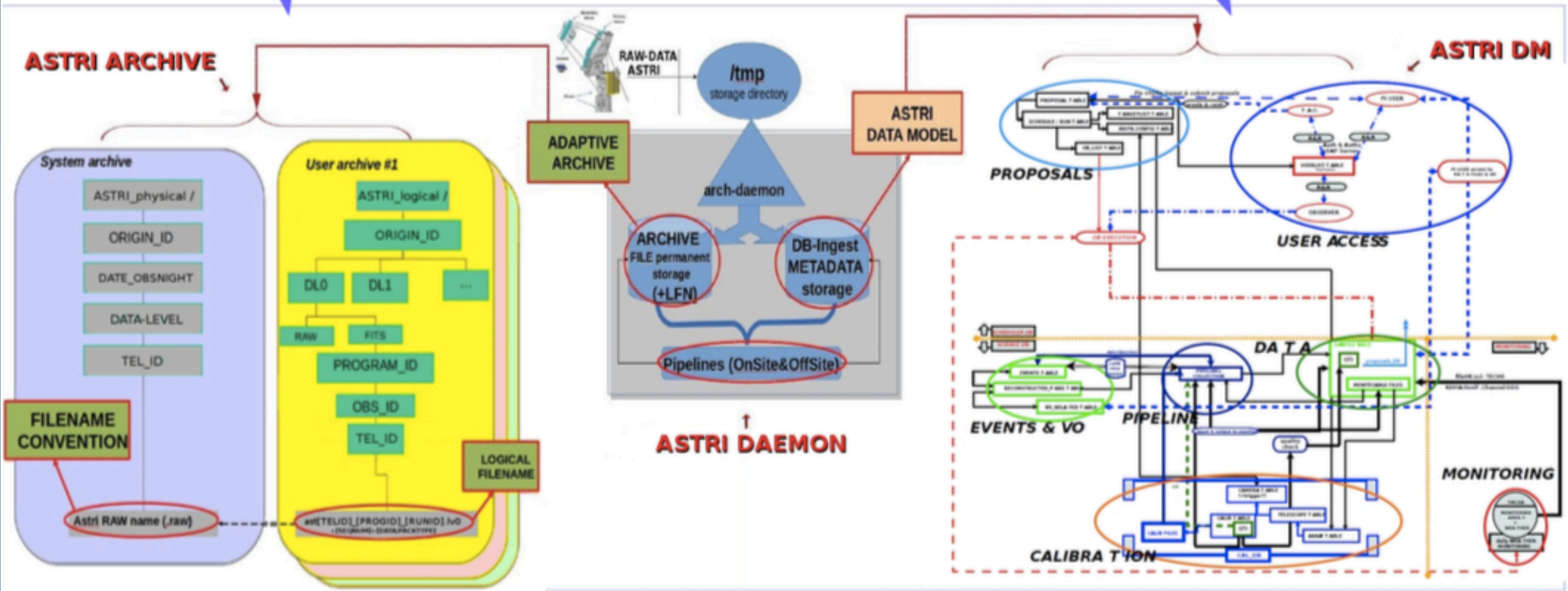
Such amounts of computing power (data reduction not included!) can be achieved with **GRID computing** / **Big Data Centers**

ASTRI Archive System

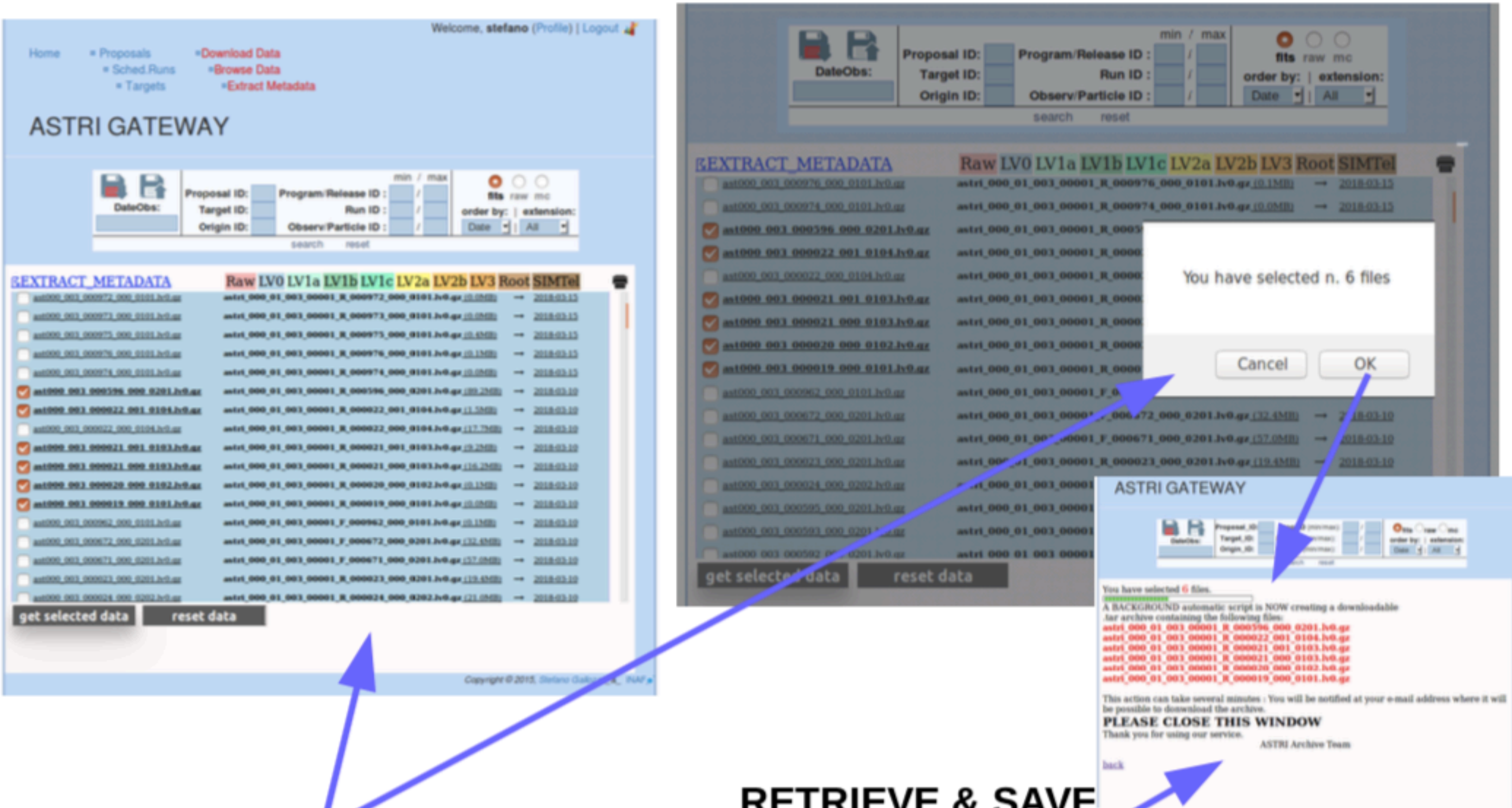
INDIGO - DataCloud

REPOSITORIES (physical & logical)

DATABASES (mySQL & mongo) + couch-base



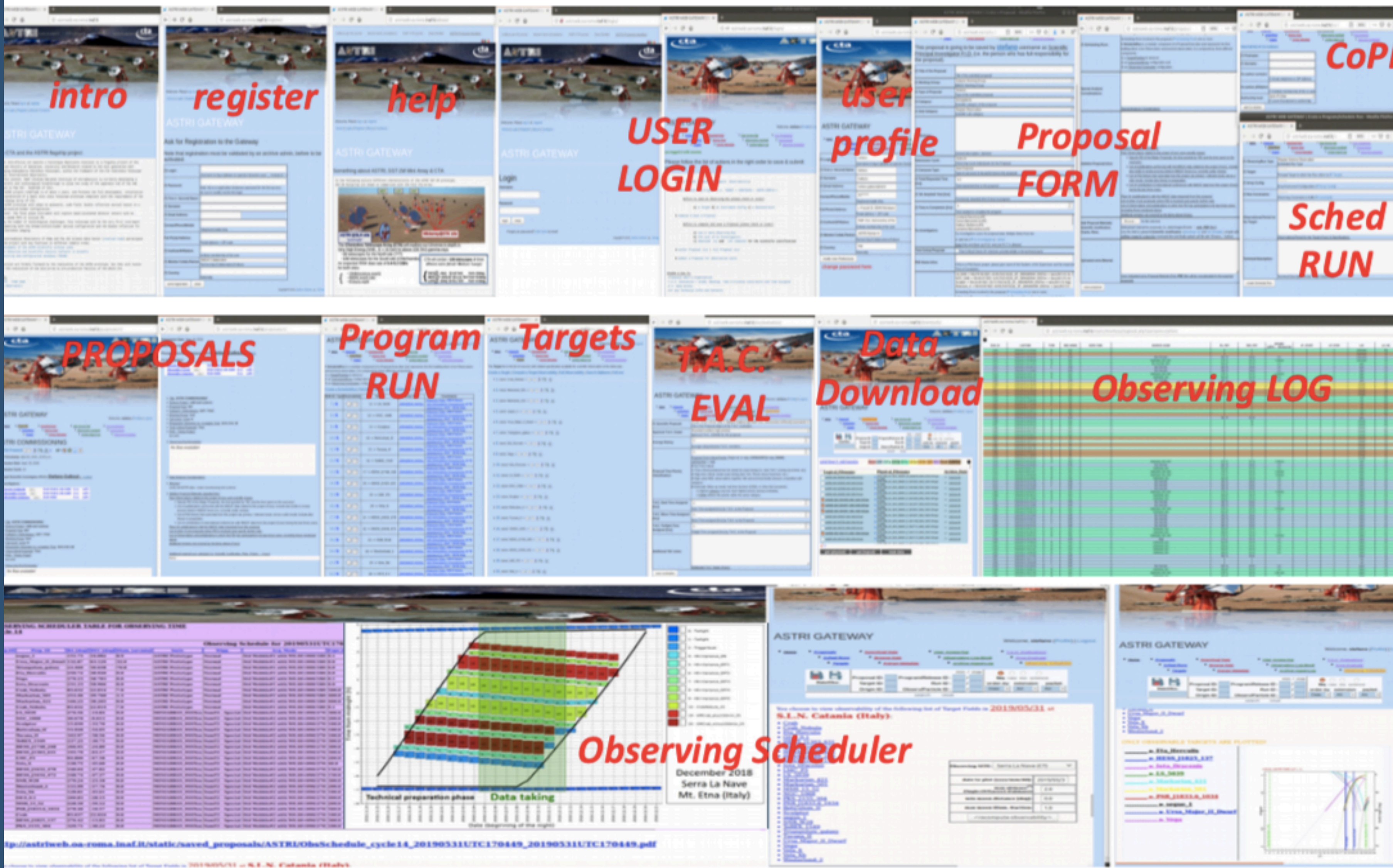
ASTRI User Access #1



QUERY & SELECT DATA (from lev0 to lev5 / photon list)

RETRIEVE & SAVE ALL LEVEL DATASETs

ASTRI Gateway



Summary and outlook

- ASTRI Data Center in Rome:
  - Efficient handling of ASTRI prototype and ACDC Project data
  - It is going to be enlarged, involving 3 distinct sites (INAF-OAR, INFN-Frascati, ASI-SSDC) in order to be ready for ASTRI mini-array data handling
- ASTRI data reduction and analysis:
  - Proper processing of ASTRI prototype real data (Crab Nebula detection!); adopted for the ACDC Project for ASTRI mini-array IRFs generation
  - Ready in its core components for the ASTRI mini-array real data reduction (further development and improvement already planned)
- ASTRI MC simulation:
  - Entire chain in operation for both ASTRI prototype and mini-array
  - MC validation activities on-going
- ASTRI Archive System and Gateway:
  - Proper archiving and retrieving of ASTRI prototype data and ACDC Project data
  - Distributed approach tested with ASTRI prototype data in view of the ASTRI mini-array application

BACKUP SLIDES

The ASTRI-Horn telescope

End-to-end prototype installed at Serra La Nave observatory (Mt. Etna, Sicily)

Mainly a technological (HW&SW) demonstrator
Telescope *verification phase*: 2017-2018
***Scientific validation phase* by 2019**

Telescope characteristics:

- **Optical design = Schwarzschild-Couder**
- Primary mirrors = 4.3 m (segmented)
- Secondary mirror = 1.8 m (monolithic)
- F/D_1 = 0.5; F = 2.15 m
- M1-M2 distance = 3.0 m
- Effective Area = 6.5 m^2

Camera properties:

- **Sensor type = SiPMs**
- Number of PDMs = 21(37)
- Number of logical pixels = 1344(2368)
- Pixel size = 0.19° (plate scale = 37.5 mm/°)
- Field of View = 7.6°(10.9°)

Expected performance:

- Energy threshold ≈ 1 TeV
- Energy/Angular resolution ≤ 25% / ≤ 0.15°
- **Sensitivity ≈ 1 Crab @ 5 σ in few hours**

The ASTRI mini-array

- **Main characteristics:**
  - **9 ASTRI-like telescopes**
    - ~250 m telescopes' relative distances
  - **Schwarzschild-Couder optical design**
    - Primary mirrors = 4.3 m (segmented)
    - Secondary mirror = 1.8 m (monolithic)
  - **SiPM sensor camera**
    - Number of logical pixels = 2368 (37 PDMs)
    - Field of View = 10.9°

- Expected performance:
  - Energy threshold ~1 TeV
  - Energy / Angular resolution ≤ 15% / ≤ 0.1°
  - **Sensitivity: better than current IACTs above ~10 TeV**

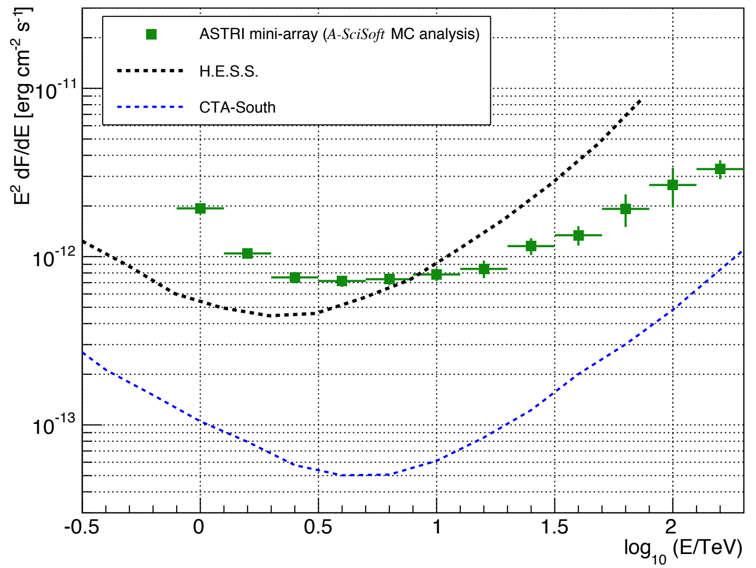
- Science cases: Galactic PWNe, SNRs, GC, bright BL Lacs and radio galaxies, extreme blazars, CR PeVatrons, Fund. Phys. and DM searches

- Synergies: LIGO/Virgo, IceCube/KM3NeT, satellites and ground-based telescopes (from radio to VHE γ-rays), …

Artistic view of the ASTRI mini-array

Credits: A. Stamerra

Expected Sensitivity of the ASTRI mini-array



Lombardi et al., SPIE 107070 (2018)

A-SciSoft general requirements

❖**CTA compliance**: *A-SciSoft* shall be as much as possible compliant with the CTA requirements and specifications and developed within the framework of the CTA pipelines sub-project;

❖**ASTRI project aim**: *A-SciSoft* shall be able to reduce data from both the ASTRI SST-2M prototype and ASTRI mini-array up to the scientific products;

❖**on/off-site processing**: *A-SciSoft* shall be able to process data both on-site and off-site;

❖**on-line processing**: *A-SciSoft* shall be able to perform an on-line data reduction during data taking in order to be able to generate real-time performance and scientific monitoring alerts;

❖**low-power consumption and parallel processing**: *A-SciSoft* shall be able to perform data reduction by means of low-power consumption and parallel computing processors (ARM/GPUs), in addition to conventional CPUs;

❖**MC data processing**: *A-SciSoft* shall be able to perform the reduction of raw MC data, in addition to real raw data;

❖**system integration**: *A-SciSoft* shall be able to efficiently interface with all external sub-systems for which an interface exists (e.g. archive system, calibration database, on-site central control software system, etc.);

A-SciSoft general requirements

- **flexibility**: *A-SciSoft* shall be flexible enough to allow an easy update in case of any possible changes in the ASTRI SST-2M prototype and mini-array hardware and raw data content/format;
- **modularity**: *A-SciSoft* shall be composed by a set of independent modules organized in efficient pipelines in order to limit inter-dependencies throughout the code and provide an easier maintainability;
- **portability**: *A-SciSoft* shall be designed for portability on UNIX-like platforms;
- **external dependencies**: *A-SciSoft* shall exclusively make use of open source libraries whose number should be minimum;
- **programming languages**: *A-SciSoft* shall be written in C++ and Python (for data processing with conventional and ARM CPUs) and CUDA (for GPU processing);
- **I/O data format**: *A-SciSoft* shall be able to handle the standard Flexible Image Transport System (FITS) data format (following the NASA-OGIP standards) for input/output (I/O) operations;
- **documentation**: *A-SciSoft* shall be extensively documented in order to allow for continual maintenance and updates;
- **high-level analysis CTA compliance**: *A-SciSoft* shall be able to generate event lists and instrumental response functions data in a format compatible with the adopted CTA Science Tools.

A-SciSoft: data levels and data types

- **EVT*n*** (*event-list data*): EVT(0,1a,1b,1c,2a,2b,3)
  - from raw data to high-level fully reduced event-list data
- **MC*n*** (*Monte Carlo event-list data*): similar definitions as in EVT*n*
- **CAL*n*** (*calibration data*): CAL(0,1,2)
  - used for cameras, optics, and array calibrations
- **MC-CAL*n*** (*Monte Carlo calibration data*): similar definitions as in CAL*n*
- **SCI-TECH*n*** (*set of technical data for scientific data reduction*): SCI-TECH(0,1,2,3)
- **LUT*n*** (*look-up-tables data*): LUT(1,2)
  - used for telescope-/array-wise event reconstruction
- **IRF*n*** (*instrument response functions data*): IRF(2,3)
- **CALDB** (*calibration database*)

FITS data format adopted

SPIE 991315 (2016); SPIE 107070 (2018)

A-SciSoft data levels

- **Level 0 (DL0)**: raw data from the hardware/software data acquisition components that are permanently archived;
- **Level 1 (DL1)**: telescope-wise reconstructed data (*reconstructed shower parameters per telescope*). Specific to ASTRI data model, the following sub-data levels are defined:
  - Level 1a (DL1a): telescope-wise calibrated data;
  - Level 1b (DL1b): telescope-wise cleaned and parameterized data (*telescope-wise image parameters*);
  - Level 1c (DL1c): telescope-wise fully reconstructed data (*telescope-wise energy, arrival direction, particle identity discrimination parameters per telescope*)
- **Level 2 (DL2)**: array-wise reconstructed data (*reconstructed shower parameters per event*). Specific to the ASTRI data model, the following sub-data levels are defined:
  - Level 2a (DL2a): array-wise merged data (*array-wise event parameters*);
  - Level 2b (DL2b): array-wise fully reconstructed data (*array-wise energy, arrival direction, particle identity discrimination parameters per event*)
- **Level 3 (DL3)**: reduced data (*selected list of events plus corresponding instrument response functions*);
- **Level 4 (DL4)**: science data (*high-level scientific data products*);
- **Level 5 (DL5)**: observatory data (*legacy observatory data and catalogs*).

SPIE 991315 (2016); SPIE 107070 (2018)

Breakdown stages and executables

Inputs: DL0
Outputs: DL1a
Modules: **astricalext;** **astrical**

aux module
pipe module

Calibration (telescope-wise)
Apply data calibration algorithms and perform data quality checks

Inputs: DL1a
Outputs: DL1c
Modules: **astricleanpar;** **astrilutsST;** **astrirecoST**

aux module
pipe module

ST=single-telescope

Inputs: DL1c
Outputs: DL2b
Modules: **astrimer;** **astriluts^{A};** **astrireco^{A}**

aux module
pipe module

A=array

Reconstruction
(telescope-wise)
Image cleaning and parameterization + single-telescope event reconstruction
(array-wise)
Stereo parameters and reconstruction (event direction, energy, classification)

Inputs: DL2b
Outputs: DL3
Modules: **astriirf;** **astriana**

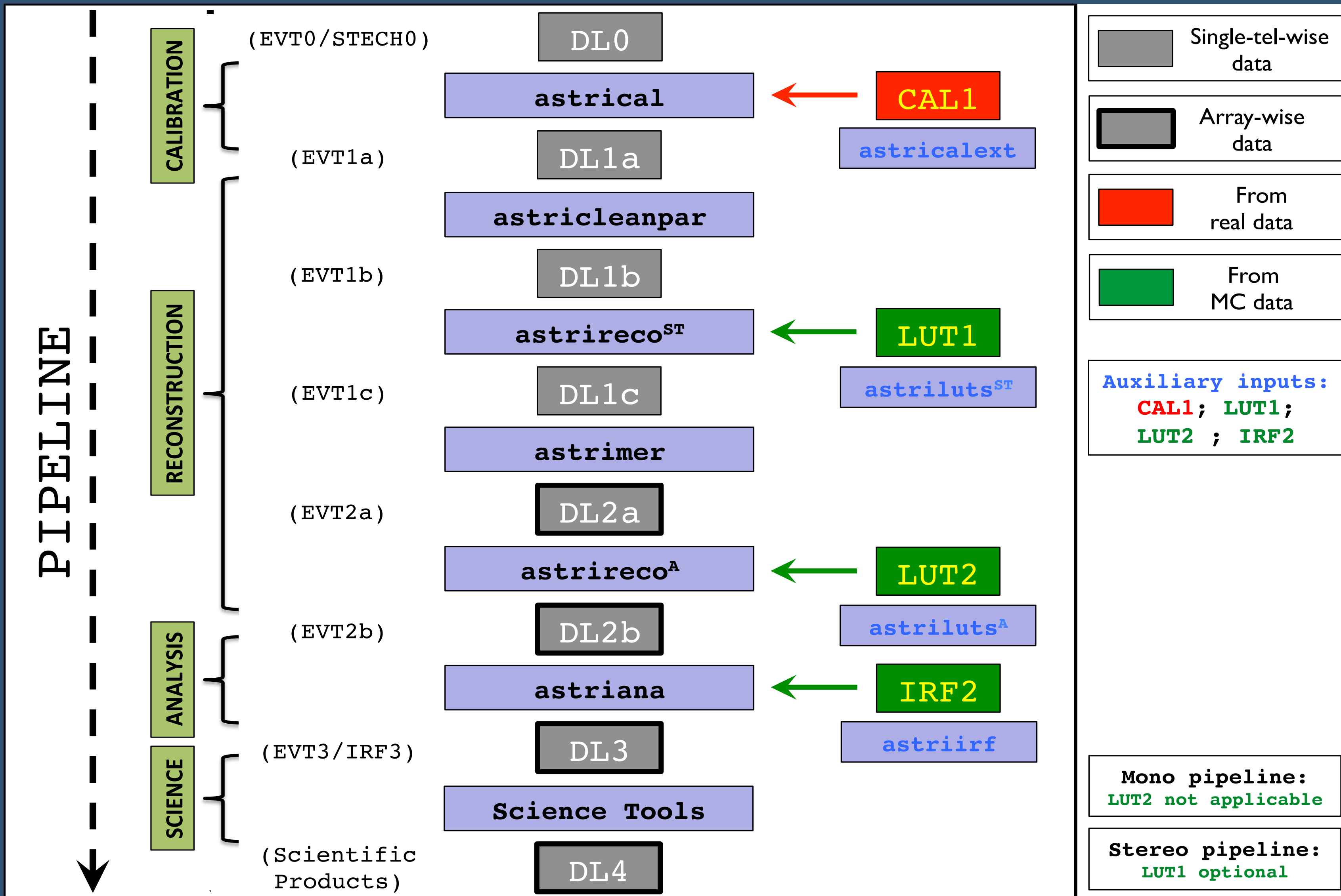
aux module
pipe module

Analysis (event-wise)
Production of gamma event list, IRFs and good time intervals (GTI)

Inputs: DL3
Outputs: DL4
Tools: **ASTRI/CTA SCIENCE TOOLS**

Science
Production of detection plots, spectra, sky maps, light curves

astripipe workflow



SPIE 991315 (2016); SPIE 107070 (2018); Astronomy & Computing (in prep.)

Breakdown stages; Basic functionalities; **Pipeline modules;** I/O data levels.

A-SciSoft deployment

A-SciSoft is deployed by means of ***conda* package manager**

conda: open source and language agnostic; available on many Linux distributions, OSX, Microsoft Windows

Widely used (ctapipe, gammapy, astropy, …)

A-SciSoft conda packages:

- ascisoft (mac-osx / linux-64)
- ascisoft-gpu (linux-64)

All dependencies are handled by the package manager

Software validation tests on *realistic* basis

- Validation tests for ASTRI prototype from “A-DC1”:
  - First validation of end-to-end single-telescope DL0➔DL4 *real-like* analysis (DL3➔DL4 achieved with *ctools*)

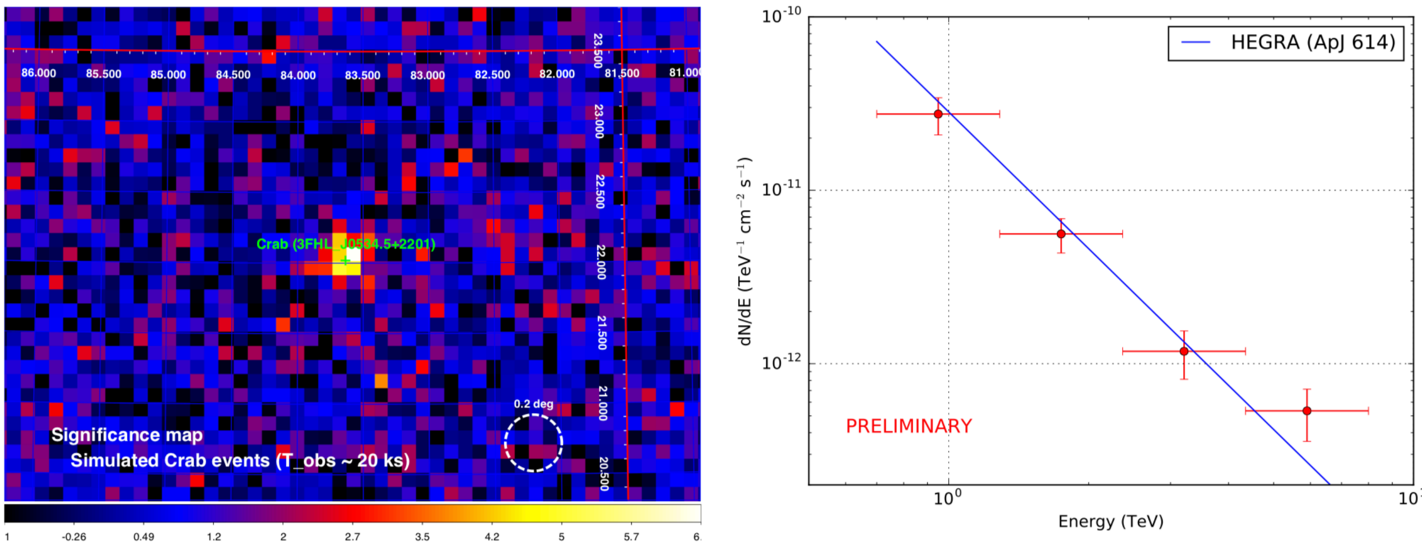


Figure 1: *Left*: Significance map of the ON data (∼5.8 hours) sky region obtained with the *ctskymap ctools* task. The white dotted circle in the lower right indicates the point-spread function (68% containment) of the analysis. *Right*: Differential spectrum of the Crab Nebula between 0.7 TeV and 8 TeV obtained from the analysis of the ON data (∼5.8 hours) obtained with the *csspec ctools* task. The blue line reprints the best-fit power-law parameterization of the Crab Nebula measured by the HEGRA Coll. [19].

PoS_ICRC2017_804 (2017) [arXiv:1709.02570v2]

ASTRI-Horn data taking: overview

- ASTRI SST-2M prototype **verification phase**:
  - **May 2017 → first light of Cherenkov events!**
  - July 2017 → first data runs (14/21 PDMs*, not nominal optics)
  - October 2017 → first scientific runs (19/21 PDMs*, not nominal optics)
  - December 2017 → scientific runs (19/21PDMs**, not nominal optics)
  - January 2018 → consistent trigger scans (21/21 PDMs**, not nominal optics)
  - May/June 2018 → first campaign on Mrk421/Mrk501 (not nominal optics)
  - **December 2018 → first extensive Crab Nebula campaign**
- Real data properly archived on-site/off-site and processed off-site (automatic on-site processing expected to be accomplished soon)
- Some HW issues (primary and secondary mirrors' reflectivity, camera HG channel, telescope pointing accuracy) → hardware improvements foreseen in the next months → nominal system condition by fall 2019
- **Scientific validation phase foreseen from fall 2019 → new Crab Nebula observations → full system characterization**

* Acquisition rate not in nominal camera condition / ** Acquisition rate in nominal camera condition

MC simulations aims

No test-beam available for IACTs!

- No way to optimize its design before completion
- No way to measure its performance (→ IRFs)

Simulations are essential through all the life of any IACT array:

1. Telescope project development
   - Design optimization
   - Layout optimization, Array trigger optimization
   - Optimization of observational strategies
2. Telescope/Array commissioning
   - Optimization of telescope performance (←→**MC validation**)
3. Data taking
   - Production of gamma samples to train gamma/hadron strategies
   - **Computation of array IRFs**
   - **Continuous MC validation**

What do we need to simulate?

1. Signal events:
   - Gamma events from point-like sources
     According to source spectrum
     According to source visibility
   - Diffuse gamma events
     Off-axis IRFs (extended sources)

2. Background events:
   - Proton events (mandatory)
   - Electron events (relevant at lowest energies)
   - Helium events
   - Heavier nuclei