Data Formats for Gamma-Ray Astronomy (GADF)

Maximilian Linhoff for the GADF Initiative

Astroparticle Symposium 2022 – Université Paris-Saclay

Astroparticle Physics – TU Dortmund | CTAO

Overview

Introduction

A (short) History of GADF

Current Status / Release 0.3

Hand-over to VODF

Open Issues

The Gamma-ray Astronomy Data Formats Initiative (GADF)

- Ground-based gamma-ray observatories are mostly run by particle physicists as closed collaborations
  - No or very limited public data access
  - Mostly proprietary analysis software
  - Custom data formats, usually strongly coupled with the proprietary software
  - Mostly based on CERN's ROOT
- Satellites and the Cherenkov Telescope Array (will) operate as open observatories
- The current generation of telescopes want a better archival format
- Strong scientific use-cases for multi-instrument analysis
- ⇒ Effort for a common, software-independent data format
- ⇒ Development of Open Source analysis software („science tools")

Imaging Air Cherenkov Telescopes

CTA Observatory / R. White (MPIK) / K. Bernlöhr (MPIK) / DESY
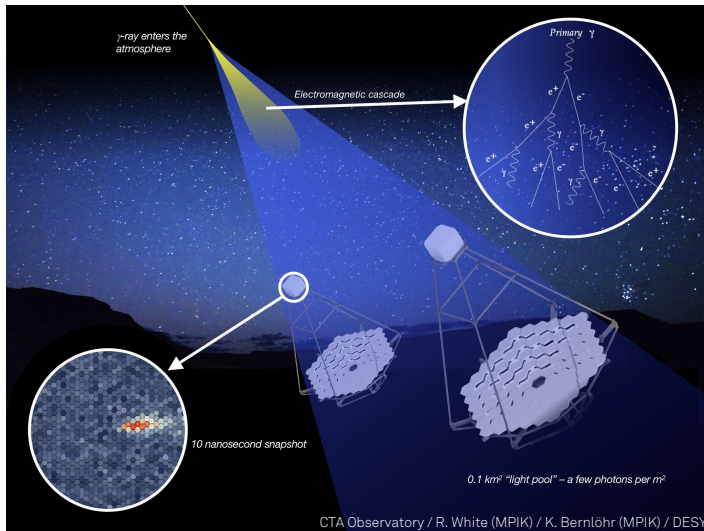
Water Cherenkov Detectors
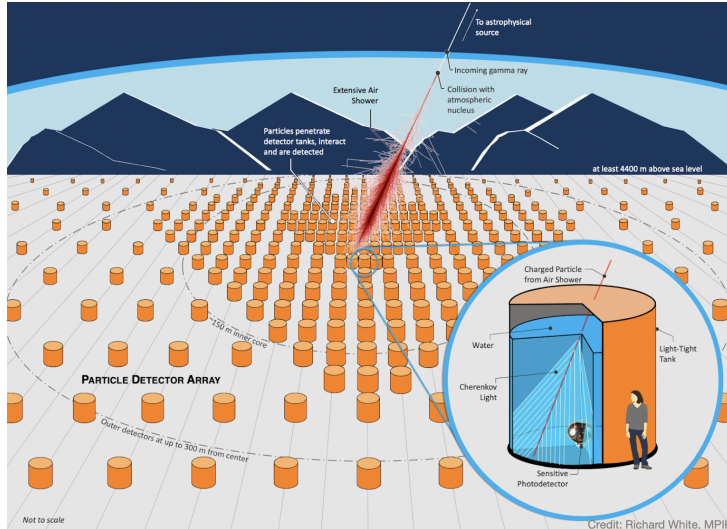
Gamma-ray Astronomy High Level Analysis

- We need to go back from the estimated quantities in the event lists to physically meaningful quantities

- Energy spectra, Light Curves, Flux Sky Maps, ...

- For this, we need the instrument response to solve the inverse problem, in general:

$$\underbrace{g(\hat{E}, \hat{\alpha}, \hat{\delta})}_{\text{Observed distribution}} = \iiint \overbrace{R(\hat{E}, \hat{\alpha}, \hat{\delta}|E, \alpha, \delta)}^{\text{Instrument Response}} \cdot \underbrace{f(E, \alpha, \delta)}_{\text{True } \gamma\text{-Signal}} \, dE \, d\Omega + \overbrace{b(\hat{E}, \hat{\alpha}, \hat{\delta})}^{\text{Background}}$$

- *R* is usually factorized into three independent components:

$$R(\hat{E}, \hat{\alpha}, \hat{\delta}|E, \alpha, \delta) = A_{\text{eff}}(E, \alpha, \delta) \cdot M(\hat{E}|E, \alpha, \delta) \cdot \text{PSF}(\hat{\alpha}, \hat{\delta}|E, \alpha, \delta)$$

- Current GADF version is mainly focused on specifying a file format for event lists and instrument response functions

A (short History of GADF)

2011 Prototypes for the CTA data format and science tools

2016
- Establishment of the Gamma-ray Astronomy Data Formats (GADF) initiative
- First preliminary release version (0.1), mainly focused on IACTs

2018
- Version 0.2 released
- Support implemented in the science tools Gammapy and ctools
- H.E.S.S. releases $\approx 50\,\text{h}$ of observations of different sources using the format

A (short History of GADF)

2019
- FACT, Fermi-LAT, H.E.S.S. MAGIC and VERITAS observations of the Crab Nebula are used to perform the first multi-instrument analysis [doi:10.1051/0004-6361/201834938]
  `https://github.com/open-gamma-ray-astro/joint-crab`
- `ctools` based analysis of the H.E.S.S. data release [doi:10.1051/0004-6361/201936010]
- Comparison of Gammapy and ctools using the H.E.S.S. data release [doi:10.1051/0004-6361/201936452]

2022
- Steering committee formed (GADF → VODF, see next talk)
- Version 0.3 released with support for wide-field instruments like HAWC
- Comparison of HAWC analysis using Gammapy + GADF data and HAWC proprietary software [doi:10.1051/0004-6361/202243527]

Current GADF Specifications

- GADF specifies a common file format for high-level gamma-ray data
- FITS with binary tables for event lists, instrument response functions, pointing information, …
- Based on the Fermi-LAT format and the OGIP standard from X-ray astronomy
- Developed on GithHub at
  `https://github.com/open-gamma-ray-astro/gamma-astro-data-formats`
- Published at `https://gamma-astro-data-formats.readthedocs.io`
- Also published and citable via Zenodo: [doi:10.5281/zenodo.1409830]

Hand-Over to VODF

- GADF had no clear organizational structure
  - No clear resolution of contentious issues
  - No clear roadmap
- Forming of a steering committee with one representative per participating instrument in 2022
- Inclusion of Neutrino observatories → renaming to VODF
- Release 0.3 of GADF planned to be last release before first VODF release

Some questions discussed but not resolved in GADF

- Which standard takes precedence? Old OGIP vs. current FITS
- Current standard is often very vague instead of descriptive
- Additional IRF parametrizations
- Uncertainties / validities of IRFs
- Multiple IRFs for one observations
- Different IRFs for different event categories („event types")
- Additional specifications for simulated datasets
- Interoperability with other entities, especially the VO