# Towards multi-instrument and reproducible gamma-ray analysis

C. Nigro* [1]

*contact:cosimo.nigro@desy.de
[1]DESY Zeuthen

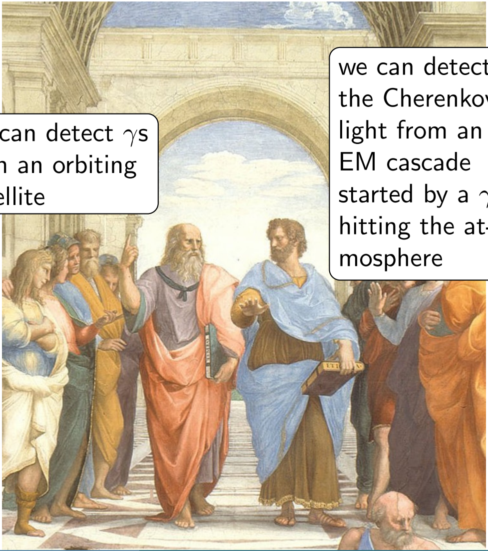
PyGamma19

18-22 March 2019, MPI for Nuclear Physics, Heidelberg

# Introduction

# The importance of a common language / format

> Different philosophies, same language.

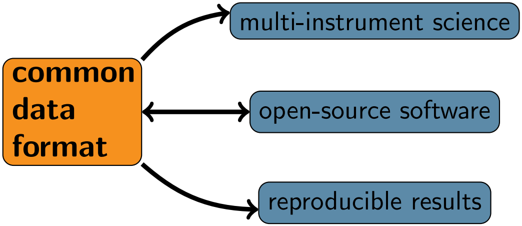


we can detect $\gamma$s with an orbiting satellite

we can detect the Cherenkov light from an EM cascade started by a $\gamma$ hitting the atmosphere

# The importance of a common language / format

- Neither data format nor software shared by gamma-ray instruments;
- current perspective, two setbacks:
  - → combination of data from different experiments needs custom expansions of proprietary analysis software;
  - → release of public legacy data needs release of analysis software;
- forward perspective, a challenge:
  - → operation of CTA as an observatory poses VHE community the problem of producing public data and analysis tools.



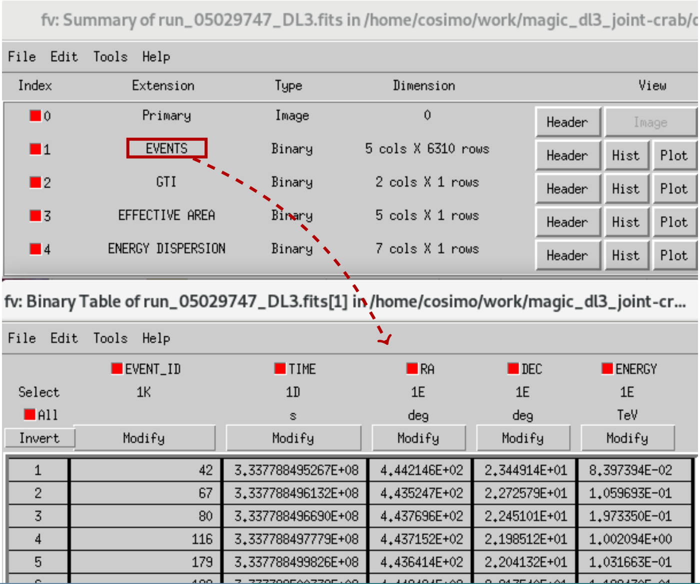
# A common gamma-ray format

- Community effort already started at *Data formats for gamma-ray astronomy* forum http://gamma-astro-data-formats.readthedocs.io;
- which level to unify?

| IACT data level | description | reduction factor |
|---|---|---|
| **DL0** | **raw** output of DAQ | |
| **DL1** | **calibrated** quantities (charge, arrival time) | 1 - 0.2 |
| **DL2** | **reconstructed** shower parameters | $10^{-1}$ |
| **DL3** | **reduced** $\gamma$ ray candidates + IRFs | $10^{-2}$ |
| **DL4** | **science** data products: spectra, LC, skymaps | $10^{-3}$ |
| **DL5** | **observatory** data: surveys, catalogues | $10^{-3}$ - $10^{-5}$ |

- space-borne instrument data (e.g. *Fermi*-LAT) can be embedded in this scheme;
- files stored in FITS format (a 30-year standard in astronomy).

# DL3 = Event Lists + IRF



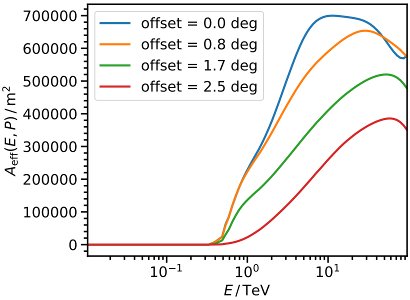

| Index | Extension | Type | Dimension | View | | |
|---|---|---|---|---|---|---|
| 0 | Primary | Image | 0 | Header | Image | |
| 1 | EVENTS | Binary | 5 cols X 6310 rows | Header | Hist | Plot |
| 2 | GTI | Binary | 2 cols X 1 rows | Header | Hist | Plot |
| 3 | EFFECTIVE AREA | Binary | 5 cols X 1 rows | Header | Hist | Plot |
| 4 | ENERGY DISPERSION | Binary | 7 cols X 1 rows | Header | Hist | Plot |



| | EVENT_ID | TIME | RA | DEC | ENERGY |
|---|---|---|---|---|---|
| Select | 1K | 1D | 1E | 1E | 1E |
| All | | s | deg | deg | TeV |
| Invert | Modify | Modify | Modify | Modify | Modify |
| 1 | 42 | 3.337788495267E+08 | 4.442146E+02 | 2.344914E+01 | 8.397394E-02 |
| 2 | 67 | 3.337788496132E+08 | 4.435247E+02 | 2.272579E+01 | 1.059693E-01 |
| 3 | 80 | 3.337788496690E+08 | 4.437696E+02 | 2.245101E+01 | 1.973350E-01 |
| 4 | 116 | 3.337788497779E+08 | 4.437152E+02 | 2.198512E+01 | 1.002094E+00 |
| 5 | 179 | 3.337788499826E+08 | 4.436414E+02 | 2.204132E+01 | 1.031663E-01 |

# DL3 = Event Lists + IRF

> Instrument Response Function (IRF) quantifies performances, transform estimated ($\hat{E}$, $\hat{\boldsymbol{P}}$) to true ($E$, $\boldsymbol{P}$) observables.



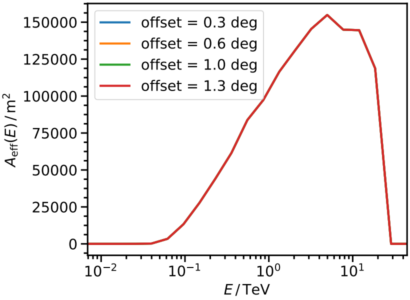

> Full-enclosure IRF:
> - → account for the dependency of the response across the FoV;
> - → current format: $\boldsymbol{P}$ = offset from camera center.

> Components:
> - → *point spread function*;
> - → *energy dispersion*;
> - → *effective area*.

> Analysis of any source in the Field of View.

# DL3 = Event Lists + IRF

> Instrument Response Function (IRF) quantifies performances, transform estimated ($\hat{E}$, $\hat{\boldsymbol{P}}$) to true ($E$, $\boldsymbol{P}$) observables.




> Point-like IRF:
  - → account for the response at the same position (offset) as the observations;
  - → $\boldsymbol{P}$ dependency removed.

> Components:
  - → no *point spread function*;
  - → *energy dispersion*;
  - → *effective area*.

> Analysis of source at fixed position (offset) in the FoV.

# The `joint-crab` effort
# arXiv: 1903.06621

# Objectives


- Using this preliminary DL3 format, we perform the first fully-reproducible multi-instrument gamma-ray analysis;
- relying on open-source software: *gammapy*;
- combining data from *Fermi*-LAT, and the four existing IACTs, to produce a joint fit of the Crab Nebula spectrum;
- online material (data and scripts): https://github.com/open-gamma-ray-astro/joint-crab/;
- DISCLAIMER: the purpose of this project is to show a method, not to provide a new measurement of the Crab Nebula spectrum.

# Datasets

- *Fermi-LAT* data freely available, IRF computed with the science tools and made DL3-compliant with `gammapy`;
- small samples of DL3 data released by IACT collaboration for this project, FACT[1] and H.E.S.S.[2] datasets already available to the public;

| Dataset | time | obs. mode | $E_{min}$ / TeV | $E_{max}$ / TeV |
|---|---|---|---|---|
| *Fermi*-LAT | $\sim$ 7 years | sky survey | 0.03 | 2 |
| MAGIC | 40 mins | pointing | 0.08 | 30 |
| VERITAS | 40 mins | pointing | 0.15 | 30 |
| FACT | 10 hours | pointing | 0.40 | 30 |
| H.E.S.S. | 3 hours | pointing | 0.50 | 30 |

- notebook to explore data: 1_data.ipynb.

[1] https://fact-project.org/data/
[2] https://www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1/

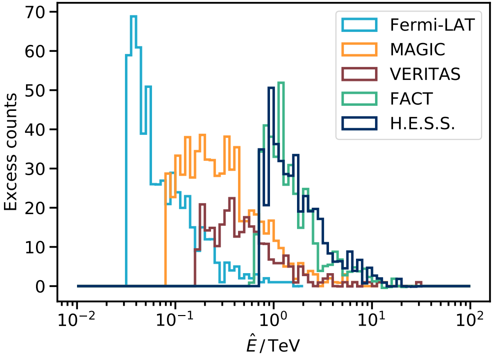
# Likelihood Analysis

- Energy spectrum $\frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\boldsymbol{\Lambda})$ estimated with a binned maximum likelihood method, $n_{\hat{E}}$ bins in estimated energy;
- data of the likelihood are counts in a signal ($N_{\mathrm{on}}$) and background ($N_{\mathrm{off}}$) sky regions, Poissonly distributed:

$$\mathcal{L}(\boldsymbol{\Lambda}|\boldsymbol{D}) = \prod_{i=1}^{n_{\mathrm{instr}}} \underbrace{\mathcal{L}_i(\boldsymbol{\Lambda}|\{N_{\mathrm{on},ijk}, N_{\mathrm{off},ijk}\}_{j=1,\dots,n_{\mathrm{runs}};k=1,\dots,n_{\hat{E}}})}_{\prod_{j=1}^{n_{\mathrm{runs}}} \prod_{k=1}^{n_{\hat{E}}} \mathrm{Pois}(g_{ijk}(\boldsymbol{\Lambda})+b_{ijk};N_{\mathrm{on},ijk})\times\mathrm{Pois}(b_{ijk}/\alpha_{ij};N_{\mathrm{off},ijk})} \quad ,$$

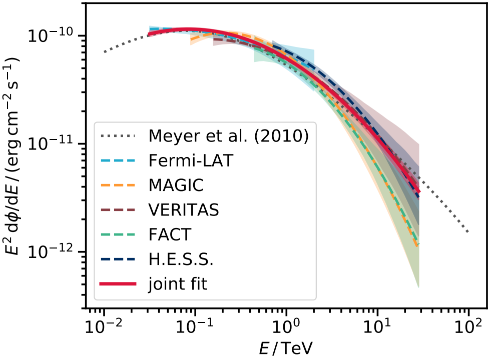
  - → $g_{ijk}$: $\frac{\mathrm{d}\phi}{\mathrm{d}E}$ folded with IRF;
  - → $b_{ijk}$: nuisance parameter, fixed at $\partial\mathcal{L}/\partial b_{ijk} = 0$;
  - → mathematical formulation in Appendix A of Piron et al. 2001;
- a joint point-like analysis is performed, *Fermi*-LAT and H.E.S.S. IRFs are reduced to a point-like format.

# Likelihood analysis




- Source counts, $N_{\rm ex} = N_{\rm on} - \alpha N_{\rm off}$, per instrument;
  - → $\alpha$ : ON to OFF exposure ratios;
- ON: circular sky regions containing the source;
- OFF: circular or ring region (free of VHE emitters) estimating the background photons to be subtracted;
- see Berge et al. 2007.

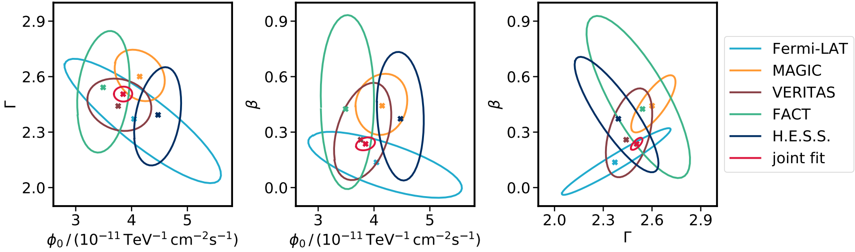
# Likelihood analysis




> Fitted spectra for log-parabolic assumed spectral model:

$$\frac{\mathrm{d}\phi}{\mathrm{d}E} = \phi_0 \left(\frac{E}{E_0}\right)^{-\Gamma - \beta \log_{10}\left(\frac{E}{E_0}\right)} .$$

# Likelihood analysis




- Likelihood contours with 68% probability content for the log-parabola parameters (individual instruments and joint fit);
- notebook to explore results: 2_results.ipynb.

# Approaches to systematics evaluation

- Identify main sources and study how their variation distort a spectral measurement (Aharonian et al. 2006, Aleksic et al. 2016):
  - → consider impact on flux normalization, spectral index and energy scale estimation;
  - → results in an additional uncertainty term, $\phi_0 \pm \sigma_{\phi_0,\mathrm{stat.}} \pm \sigma_{\phi_0,\mathrm{syst.}}$.
- Incorporate uncertainties in the likelihood (as Dickinson et al. 2013 for background subtraction uncertainty):
  - → results in a global stat + syst. uncertainty term, $\phi_0 \pm \sigma_{\phi_0,\mathrm{stat.+syst.}}$.

# Modified Likelihood

- Example of how to include the systematic uncertainties on the energy scale of the different instruments (following Dembinski et al. 2017):
  - constant energy bias per instrument $z_i = \frac{\tilde{E} - E}{E} = \frac{\tilde{E}}{E} - 1$;
  - modified assumed spectrum $\tilde{E}$:

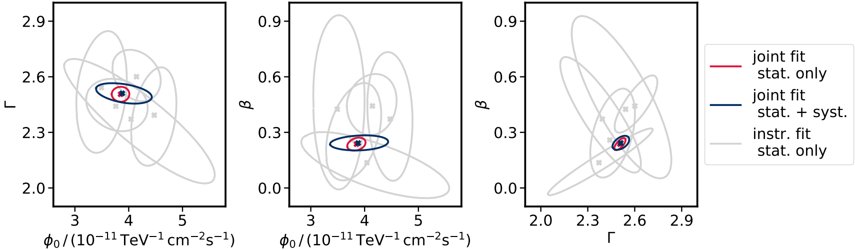
$$\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}\tilde{E}} = \frac{\mathrm{d}\phi}{\mathrm{d}E}\frac{\mathrm{d}E}{\mathrm{d}\tilde{E}} = \phi_0 \left(\frac{E/(1+z_i)}{E_0}\right)^{-\Gamma + \beta \log_{10}\left(\frac{E/(1+z_i)}{E_0}\right)} \left(\frac{1}{1+z_i}\right)$$

- global likelihood function extended with the distributions of the $z_i$:

$$\mathcal{L}(\mathbf{\Lambda}|\boldsymbol{D}) = \prod_{i=1}^{n_{\mathrm{instr}}} \mathcal{L}_i(\mathbf{\Lambda}|\boldsymbol{\mathcal{D}_i}) \times \mathcal{N}(z_i; 0, \delta_i^2);$$

  - $z_i$ fitted with the other spectral parameters;
  - constrained with $\delta_i$ = systematic uncertainty on the energy scale estimated by each instrument.

# Modified likelihood




- Likelihood contours with 68% probability content incorporating stat. and syst. uncertainties;
- notebook to reproduce fit with systematics: 3_systematics.ipynb.

# Using a theoretical model

- An analytical function is not the only possibility to perform a likelihood fit, any theoretical model can be used for $\frac{d\phi}{dE}$.
- Typically theoretical models are not plugged in the likelihood estimation but fitted to spectral points:
  - → often not unfolded (i.e. in $E_{\mathrm{est}}$), and limited in cases where the energy dispersion plays a major role.
- Releasing the results of the data reduction (i.e. ON + OFF distributions and IRF) would allow successive likelihood fit with any arbitrary theoretical model;
- example with `naima` radiative model: 4_naima.ipynb.

# How is reproducibility achieved?

- Short-term:
  - all the code will be publicly available in `GitHub` https://github.com/open-gamma-ray-astro/joint-crab;
  - the size of the data is $\sim$ MB, can be provided along with the code;
  - packages managed via `anaconda` environment.
- Medium-term:
  - it may happen that the `conda` virtual environment is not enough to guarantee reproducibility (software not anymore mantained), a joint-crab Docker container is provided.
- Long-term:
  - on-line material available on Zenodo, `DOI: 10.5281/zenodo.2381863`.

# Wrap-up and prospects

# Status

- With the current DL3 data only joint point-like analysis can be performed (among IACTs, only H.E.S.S. released full-enclosure IRF for this project);
- see Gernot, Lars and Lea's talk for an update of the DL3 converter status in VERITAS, H.E.S.S. and MAGIC;
- the format itself needs to be extended: expressing the $\boldsymbol{P}$ dependency as a radial offset is not enough for experiments with non-symmetric camera acceptances (e.g. MAGIC).

# An open gamma-ray science

- What can our community achieve?
- An approach to gamma-ray science, summarized by three essential concepts: common data format, open-source software and reproducible results, the first being the cornerstone of the last two.
- With the `joint-crab` example we illustrate **this approach is already within our reach**!
- A key asset for future gamma-ray instruments like CTA that will be operated as an open observatory and share its data with a wide astronomical community.