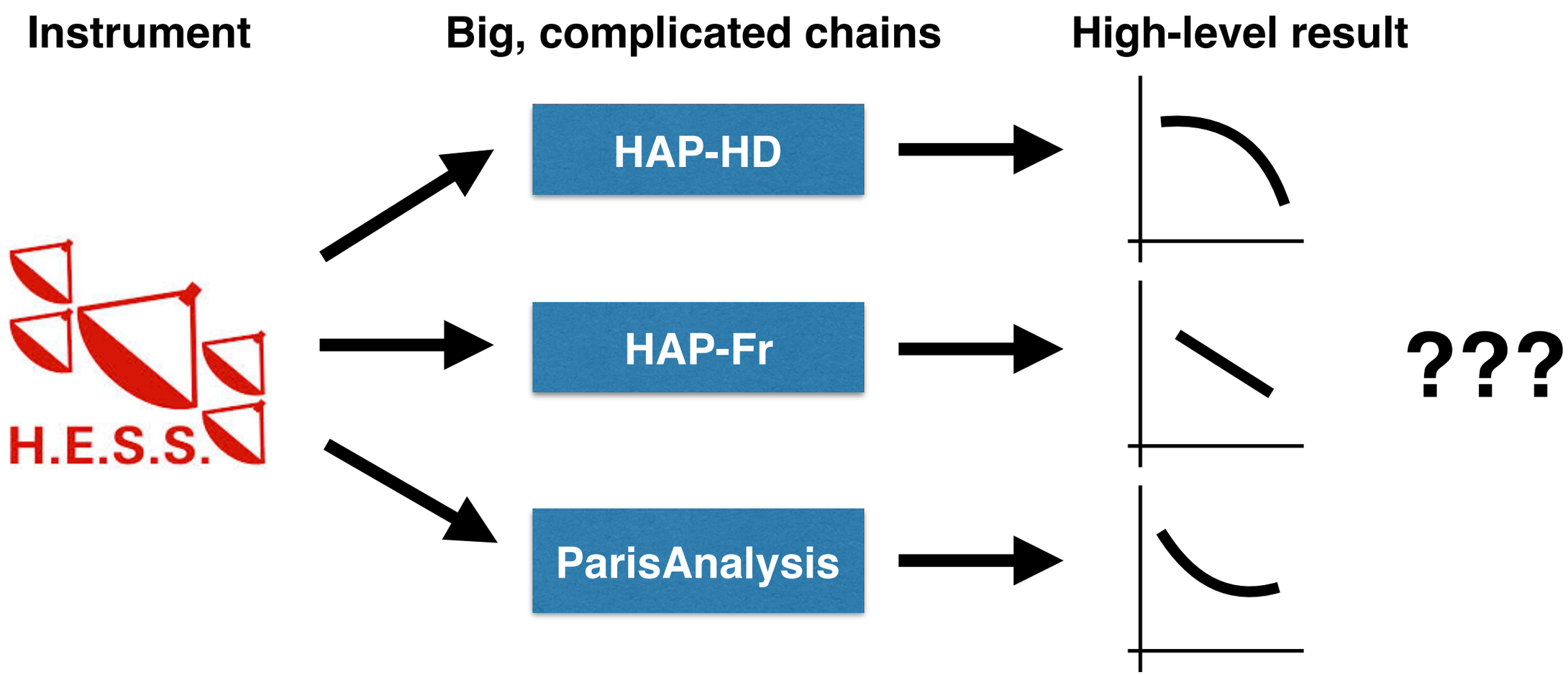
DL3 IN H.E.S.S.

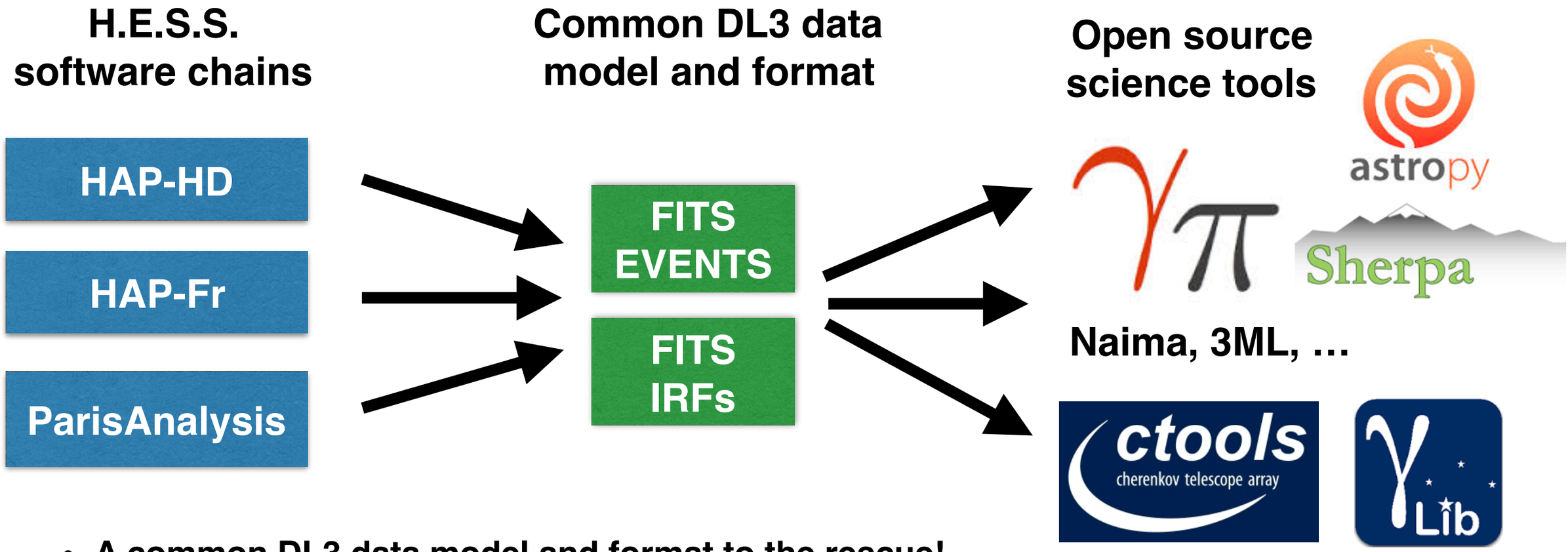
Christoph Deil, MPIK Heidelberg
IACT DL3 meeting, Meudon, April 6, 2016

H.E.S.S. data analysis with internal chains



- In H.E.S.S. we have 3 internal analysis chains (HAP-HD and HAP-Fr share some parts)
- DL1 to DL5 levels are mostly incompatible between the chains
- Lack of common data formats prevents mix & match of methods from different chains (e.g. calibrations, reconstructions, gamma-hadron separations, high-level analysis) and makes it hard to understand the differences in high-level results (e.g. source spectrum).

H.E.S.S. data analysis with open source tools



- **A common DL3 data model and format to the rescue!**
- Export events and IRFs at the DL3 level from each chain (after gamma-hadron separation, similar to Fermi-LAT public data)
- Makes mid-level (event energies, positions) and high-level (source position, morphology, spectrum) checks between the different chains, algorithms and open-source tools possible.
- Work on data formats, exporters, checks, open source tools ongoing in parallel …

Data format summary

H.E.S.S. software chains

HAP-HD

HAP-Fr

ParisAnalysis

FITS exporter

Per observation (valid for all targets in the field of view) → Per target (spectral analysis only)

EVENTS header
- Start time
- Observation time
- Livetime
- Pointing position (RA / DEC)

EVENTS table
- Time
- Energy
- RA, DEC
- DETX, DETY

Instrument response functions (IRFs) in Fermi-LAT IRF format

- Effective area (offset, energy)
- Energy resolution (offset, energy, energy_reco/energy)
- Point spread function (offset, energy)
- Background (detx, dety, energy_reco)

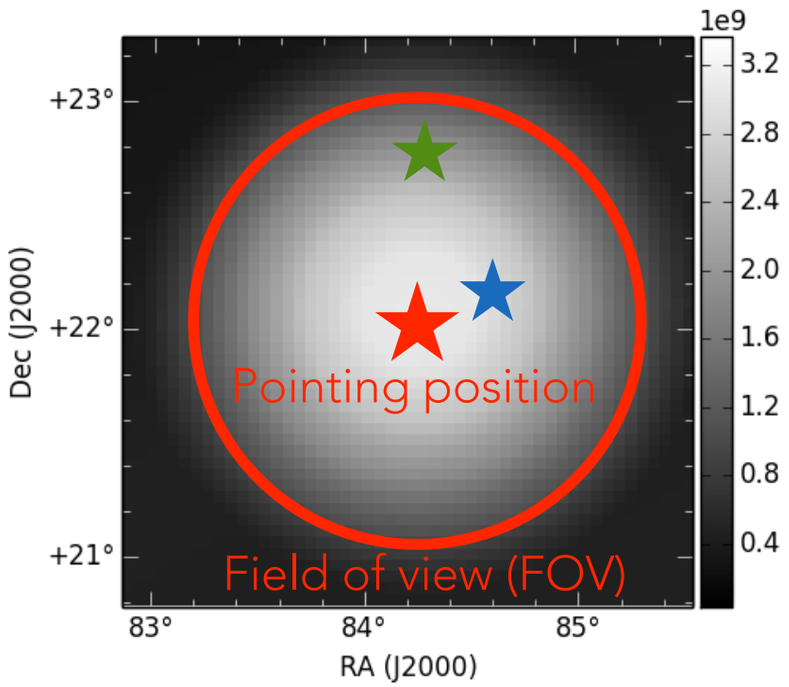
OGIP standard format

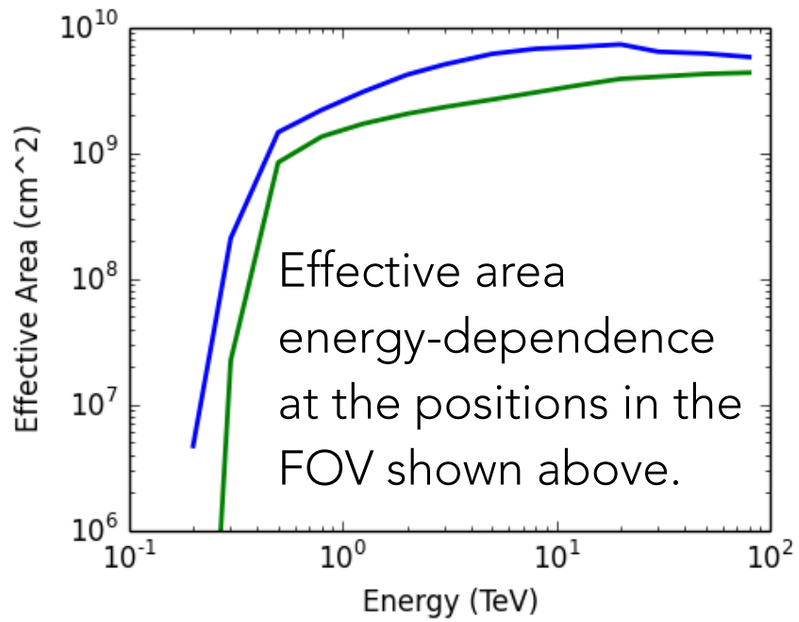
- Effective area — "ARF format" (energy)
- Energy resolution — "RMF format" (energy, energy_reco)

PSF formats "RPSF" and "REEF" not exported and used so far

- ON and OFF region counts (energy) — "PHA format"

- Mostly agreed on FITS formats for EVENTS, AEFF, EDISP, PSF, BACKGROUND in H.E.S.S. and open source tools (Gammapy & Gammalib)
- One exporter implemented in each H.E.S.S.-internal chain, export data once for all observations (per config), valid for the whole field of view (FOV)
- To analyse any target, use response for a given FOV position (or model the whole FOV), in any case no need to go back to the H.E.S.S software for any high-level analysis.



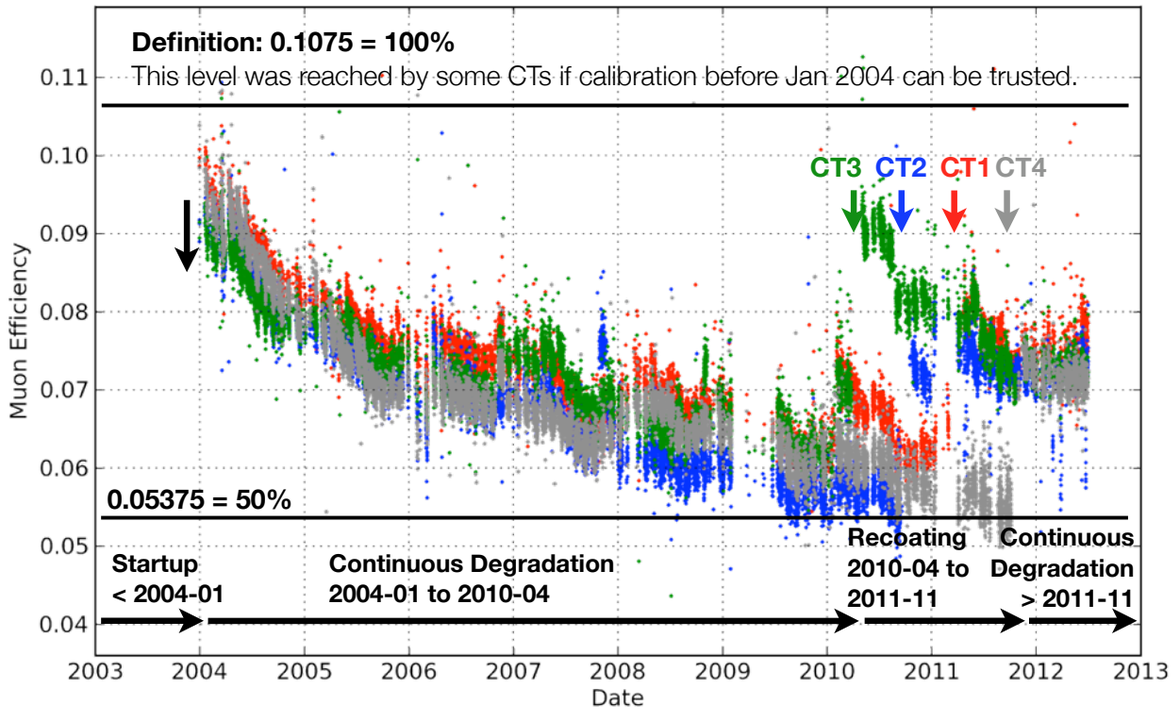


Why export per-observation event lists / IRFs?

- The whole H.E.S.S. 1 dataset is ~ 1 GB, fits on a USB stick and in memory -> super convenient and fast high-level analysis.
- Contrast this to the H.E.S.S. software, which uses a global IRF lookup database (ROOT files, sometimes MySQL database access for tech data like muon efficiency or quality selection).
  Global IRF database:
  - Much larger (10+ GB) because of an ever-changing detector (many different array configurations, optical efficiency degradation, mirror re-coatings, new telescopes and cameras).
  - Very complex, MC productions and IRF storage handled differently in each chain.
- **Pre-computed per-observation IRFs are awesome for end users and science tool writers!**

Muon Efficiency 2004 – 2012

CT1, CT2, CT3, CT4 (HD calibration, run-by-run)



IRF GLOBAL DATABASE DEPENDENCIES

lookup file (.root)	histogram	parameters	x-axis	y-axis	z-axis
ScaleInfo	avg_length	opt,azm,zen,off	$\ln(\text{size}/p.e.)$	d/m	$\langle L \rangle$/mrad
	avg_width	opt,azm,zen,off	$\ln(\text{size}/p.e.)$	d/m	$\langle W \rangle$/mrad
	sigma_length	opt,azm,zen,off	$\ln(\text{size}/p.e.)$	d/m	σ_L/mrad
	sigma_width	opt,azm,zen,off	$\ln(\text{size}/p.e.)$	d/m	σ_W/mrad
EnergyInfo	MeanTrueEnergy	opt,azm,zen,off,tel	$\ln(\text{size}/p.e.)$	d/m	E/TeV
	SigmaTrueEnergy	opt,azm,zen,off,tel	$\ln(\text{size}/p.e.)$	d/m	$\sigma(E)$/TeV
EffectiveAreas	EffArea_TrueEnergy	opt,telp,azm,zen,off	E_{true}/TeV	$A_{\text{eff}}/\text{m}^2$	
	EffArea_RecoEnergy	opt,telp,azm,zen,off	E_{reco}/TeV	$A_{\text{eff}}/\text{m}^2$	
	EnergyBias	opt,telp,azm,zen,off	$\log_{10}(E/\text{TeV})$	$(E_{\text{reco}} - E_{\text{true}})/E_{\text{true}}$	
PSF	ThetaSq	opt,telp,azm,zen,off	$\log_{10}(E/\text{TeV})$	θ^2/deg^2	p.d.f. value
EnergyReconstruction	EnergyReconstructionPDF	opt,telp,azm,zen,off	$\log_{10}(E/\text{TeV})$	$(E_{\text{reco}} - E_{\text{true}})/E_{\text{true}}$	p.d.f. value
RadialAcceptance	RadialLookup	zen	$(\Delta\Psi)^2/\text{deg}^2$	acc/a.u.	

Table 1: Lookups used in `hap`. Here, d is the impact distance, L and W are the length and width of the shower in the camera, respectively. $(\Delta\Psi)^2$ is the square of the angular distance to the observation position. θ^2 is the square of the angular distance to the centre of a source. Azm, zen, off, opt,tel, and telp are used to abbreviate azimuth, zenith and offset angles, optical efficiencies, telescope ID, and telescope pattern (eq. (5)), respectively.

DL3 data distribution in HESS

Master index JSON file

```
{
  "datasets": [
    {
      "name": "hd-hap-prod1-stdcuts",
      "obsindx": "relative/path/to/obs-index.fits.gz",
      "hduindx": "relative/path/to/hdu-index.fits.gz",
    },
    {
      "name": "pa-prod3-hardcuts",
      "obsindx": "relative/path/to/obs-index.fits.gz",
      "hduindx": "relative/path/to/hdu-index.fits.gz",
    }
    ]
}
```

- See IACT data storage spec
- No DL3 TECH files!

HDU index FITS table columns

Column Name	Description
`OBS_ID`	Observation ID (a.k.a. run number)
`HDU_TYPE`	HDU type (see below)
`HDU_CLASS`	HDU class (see below)
`FILE_DIR`	Directory of file (rel. to this file)
`FILE_NAME`	Name of file
`HDU_NAME`	Name of HDU in file

Valid `HDU_TYPE` values (others opt

- `events` - Event list
- `gti` - Good time interval
- `aeff` - Effective area
- `psf` - Point spread function
- `edisp` - Energy dispersion
- `bkg` - Background

Valid `HDU_CLASS` values:

- `events` - see format spec: IACT event lists
- `gti` - see format spec: TODO
- `aeff_2d` - see format spec: aeff_2d format
- `edisp_2d` - see format spec: edisp_2d format
- `psf_table` - see format spec: psf_table format
- `psf_3gauss` - see format spec: psf_3gauss format
- `psf_king` - see format spec: psf_king format
- `psf_gtpsf` – see format spec: gtpsf format
- `bkg_2d` - see format spec: bkg_2d format
- `bkg_3d` - see format spec: bkg_3d format

HESS EXPORTED FITS FILES

```
.
├── background
│   ├── bgmodel_alt7_az0.fits.gz
│   └── bgmodel_alt8_az0.fits.gz
├── hdu-index.fits.gz
├── obs-index.fits.gz
└── run23400-23599
    ├── run23523
    │   ├── aeff_2d_23523.fits.gz
    │   ├── edisp_2d_23523.fits.gz
    │   └── psf_king_23523.fits.gz
    ├── run23526
    │   ├── aeff_2d_23526.fits.gz
    │   ├── edisp_2d_23526.fits.gz
    │   └── psf_king_23526.fits.gz
    ├── run23559
    │   ├── aeff_2d_23559.fits.gz
    │   ├── edisp_2d_23559.fits.gz
    │   └── psf_king_23559.fits.gz
    └── run23592
        ├── aeff_2d_23592.fits.gz
        ├── edisp_2d_23592.fits.gz
        └── psf_king_23592.fits.gz
```

- Formats described in the open-astro-gamma-data spec (see previous presentation.
- On the left an example:
  files for 4 Crab runs from HESS.
- 1 observation = 1 GTI = 28 min
- Pre-computed IRFs
- IRF association currently via OBS_ID
- Obs index table for quick data selection and HDU index table for quick data localisation (HESS 1 has 20k obs and 80k HDUs)
- Background models are shared between many observations (~ 10 to 100 bg models only)

HESS SHORT-TERM PLAN (WEEKS)

- Improve open data spec
  - FOV coordinates
  - Change IRF - EVENT association via GTI or RTI?
- Improve exporters
  - Which PSF to use? EDISP and background smoothing?
  - Adapt to FOV and GIT/RTI changes in the spec

HESS MID-TERM PLAN (MONTHS)

- Extend open data spec to support splitting the data in event classes for HESS 2 mono / stereo analysis (and other classes):
  - Introduce EVENT_TYPE
  - Change EVENT - IRF association method:
    one set of IRFs per GTI and EVENT_TYPE.
- We are preparing a proposal for a public HESS test data release (~50 hours of decade-old HESS 1 data, two point sources, two extended sources, one variable source).
  Goal: contribution to data model / format / science tool writers, not a science data release.
- More testing of exporters and tools (X-check task)
  (Also: use for science publications!)

BACKUP SLIDES

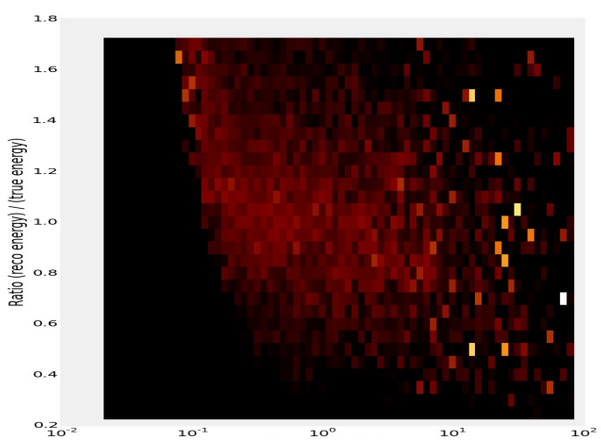
Issues with histogram IRFs

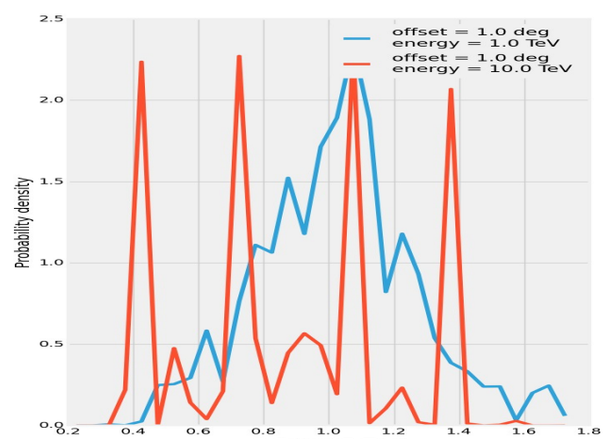
- In HESS we use histogram IRFs for energy dispersion (edisp2d) and background (bkg3d).

- As you can see on the right, these can become very noisy…

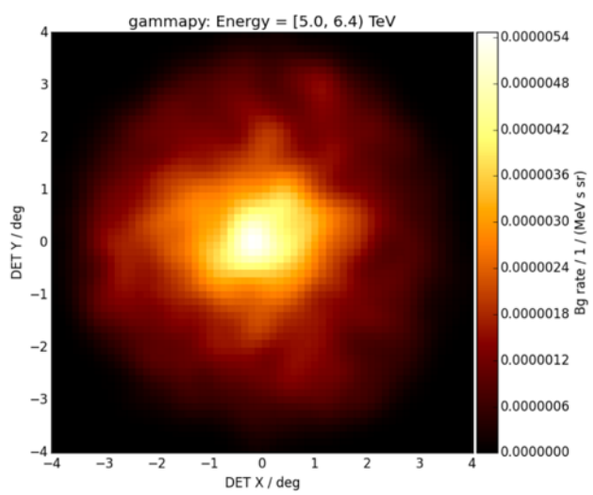
- Started investigating smoothing and analytical models for energy dispersion and background …

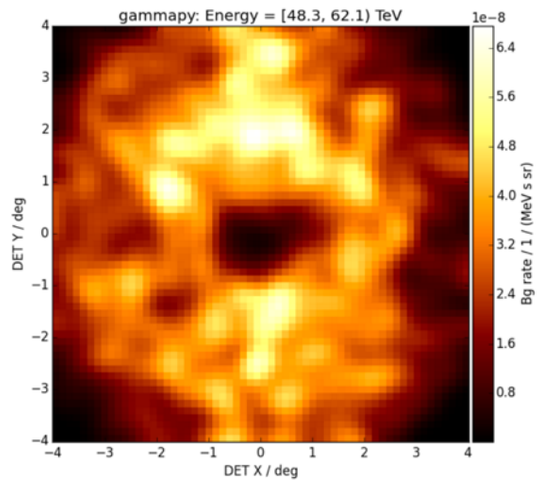
Energy dispersion example





Background example





Low energy: stats OK!

High energy: low stats!

Issues with analytical IRFs

- In HESS we use analytical PSF models, such as _psf_3gauss.

- For some configs / observations / energy / offset bins, the analytical fit doesn't converge.

- Started investigating if 2-Gauss or King profile is good enough, or if smoothed histogram PSF would work better.

Incorrect PSF model because of non-converging triple-Gauss PSF analytical model fit.