Software and data preservation for the VERITAS gamma-ray observatory

MU Days 2024 HH

Gernot Maier

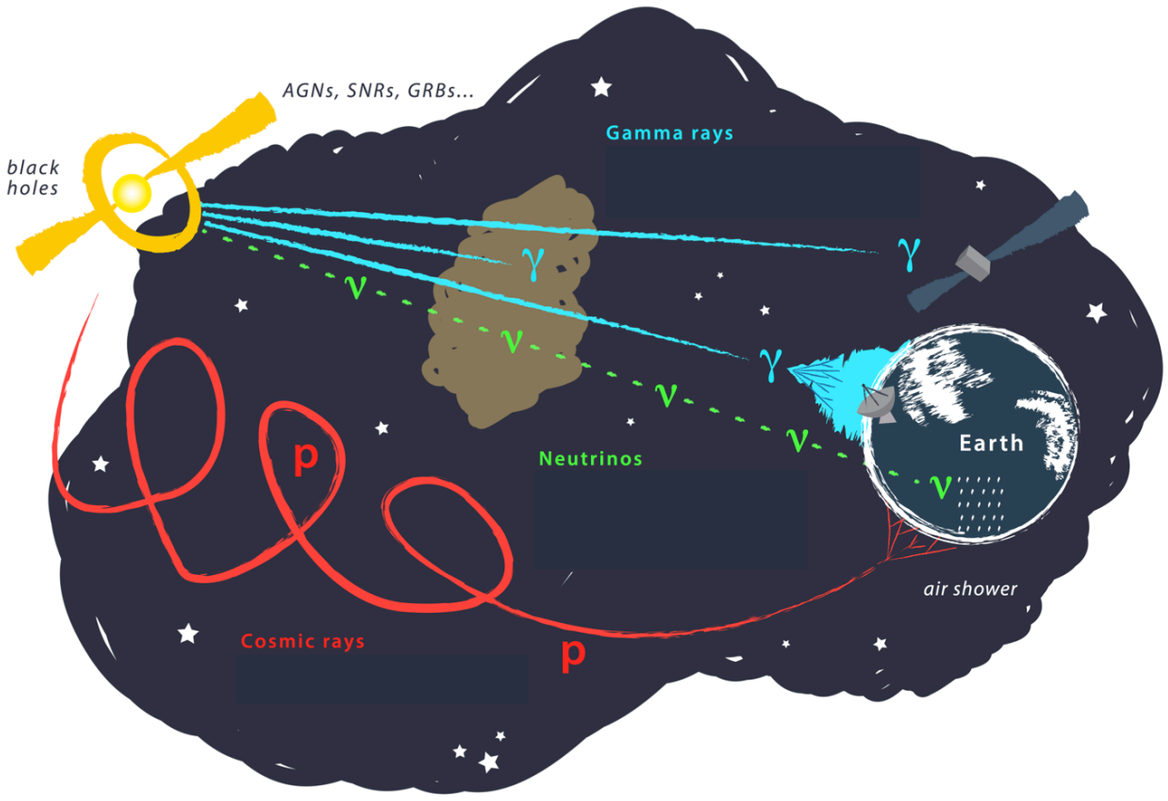
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

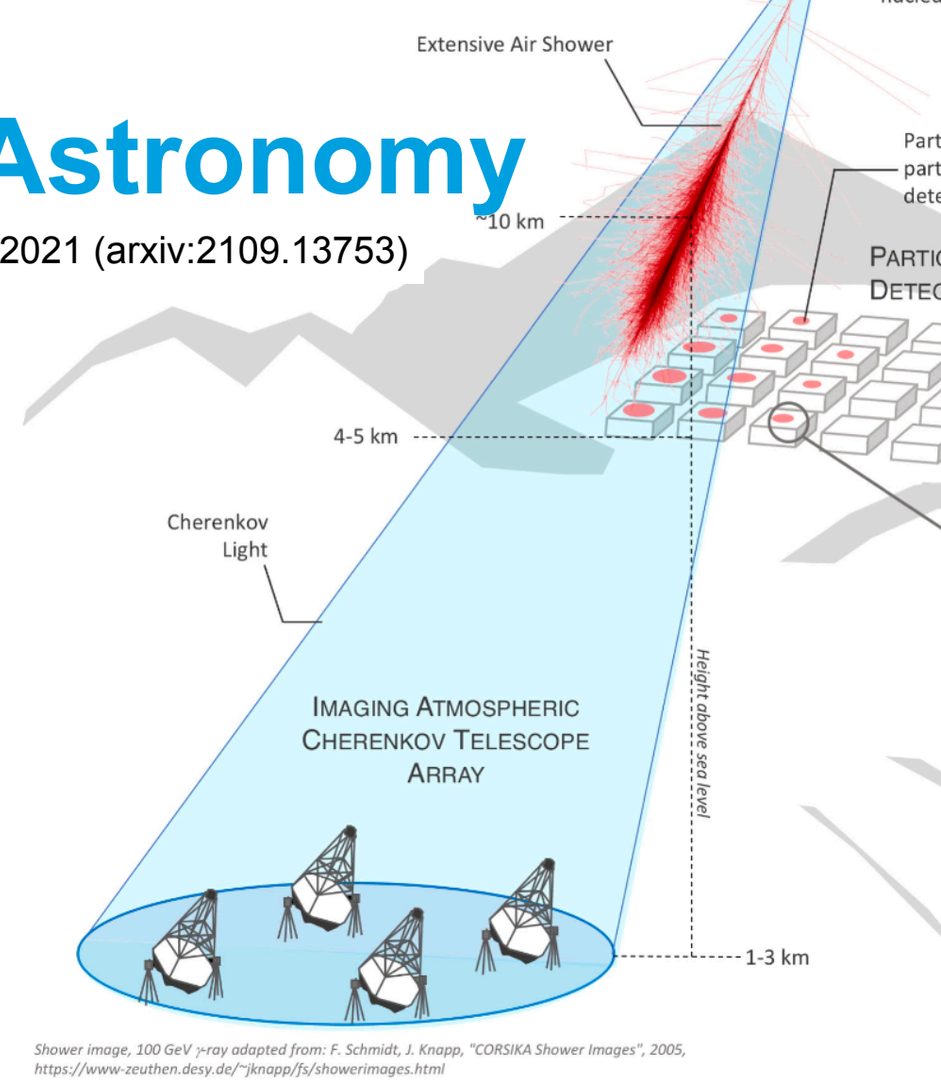
Analyse- und Datenzentrum für die Multi-Messenger Astroteilchenphysik

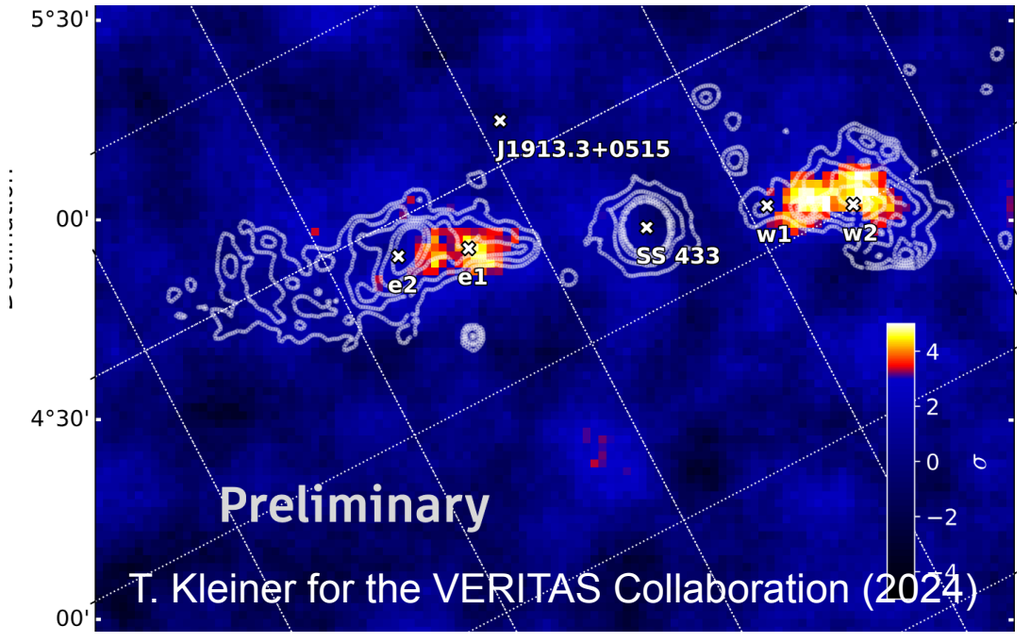
ADC-MAPP

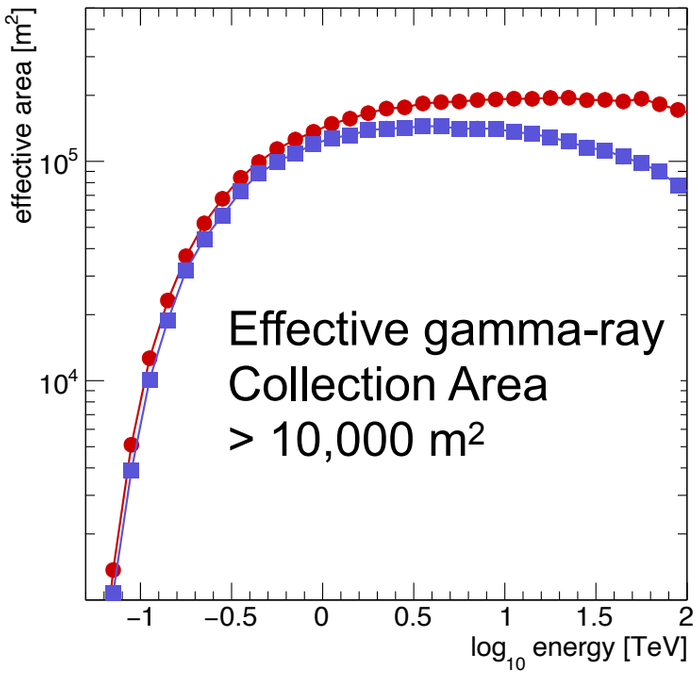
Very-high energy Gamma-ray Astronomy

Mitchell 2021 (arxiv:2109.13753)









The VERITAS Observatory

The Very Energetic Radiation Imaging Telescope Array System

- Location @ Amado, AZ
- Array of four 12-m imaging air Cherenkov telescopes (3.5° f.o.v)
- Sensitive 85 GeV to > 30 TeV
- PSF 0.08° @ 1 TeV
- 1% Crab in < 25 hrs

VERITAS Collaboration

Facility and Collaboration:

- Located at the Fred Lawrence Whipple Observatory in southern Arizona @ 1,268 m a.s.l.
  - CTA prototype SCT telescope co-located
- Starting 18th year of full-array operations this month.
- International Collaboration:
  - ~80 members incl. 20 graduate students and 8 Postdocs
  - +10 active Associate Members

Funding Sources:

- **USA**: National Science Foundation, Smithsonian Astrophysical Observatory
- **Canada**: Natural Sciences and Engineering Research Council
- **Germany**: Helmholtz Association

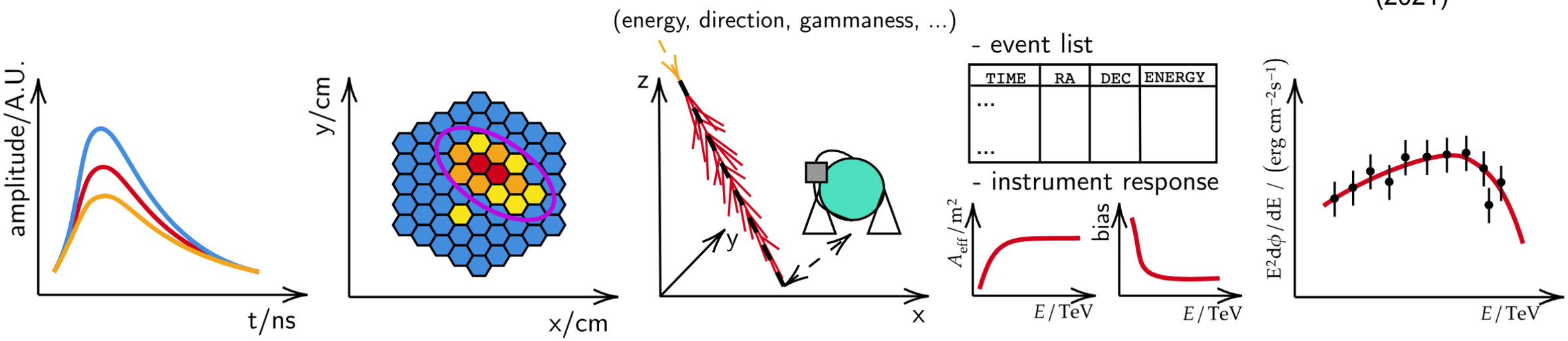
News:

- Plans to apply for operations support from NSF for 2025-2028 window.

In operation since 2007: collected more than 20,000 hours of observations
How to ensure long-term access to VERITAS data?

VERITAS Data levels and data products

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)

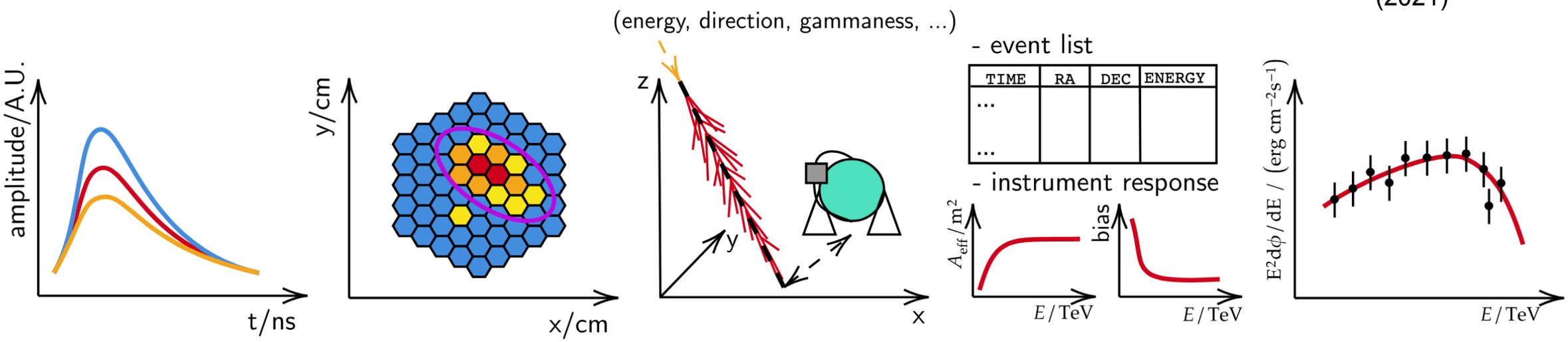


DL0 raw DAQ output → **DL1** image parameters → **DL2** shower parameters → **DL3** event list + IRF → **DL4** science products

- Ideally all data and software should be public
  (but infrastructure, collaboration effort, property rights, funding of data preparation)
- Data and software interlinked and hard to separated.
- Data processing chain developed in 2003-2005
- **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

VERITAS DL0 - Raw data.

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)



VERITAS Simulation Pipeline

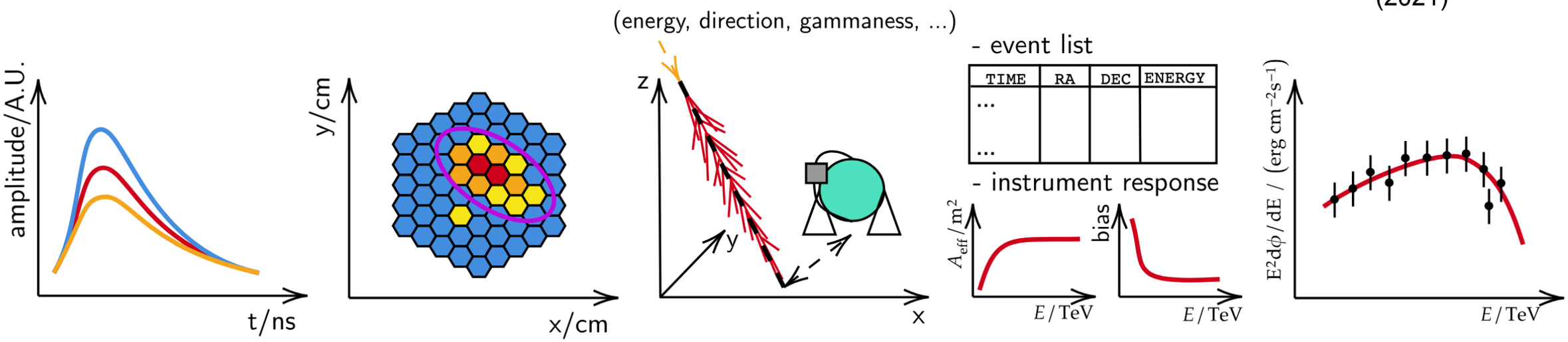
DOI 10.5281/zenodo.13943452 License BSD 3-Clause build-images

- Proprietary data format (VBF) *internal*
- MC Simulations: software_plus containers *public*
- Calibration/instrument configuration: mySQL DB *internal*

- Data: 1.5 PB at UCLA; simulations: 500 TB at DESY/NERSC *internal*

VERITAS DL1 - Images

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)



DL0 raw DAQ output → **DL1** image parameters → **DL2** shower parameters → **DL3** event list + IRF → **DL4** science products

- Calibrated images and image parameters (ROOT-format)
- Reconstruction code Eventdisplay *public*
  - Code, containers, scripts, workflows, configuration files

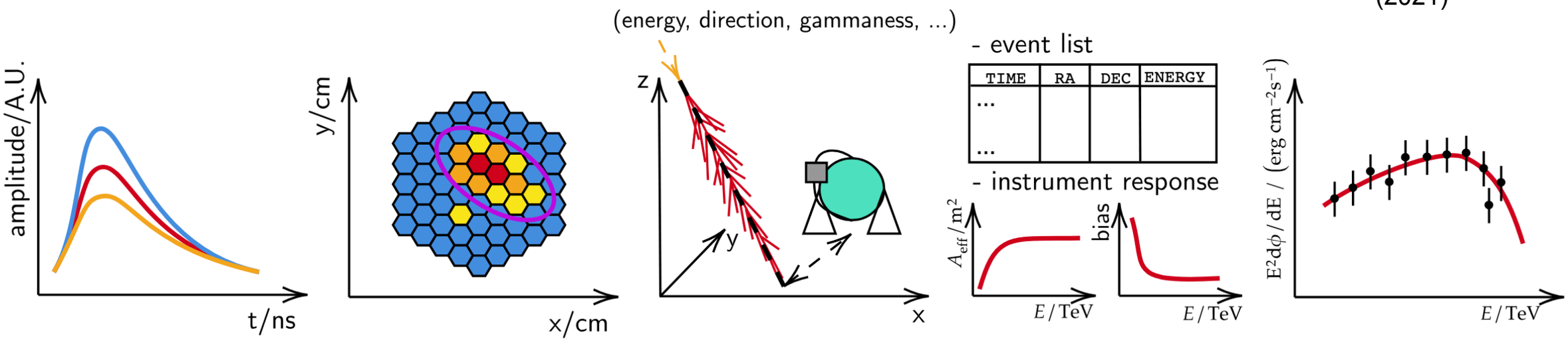
Eventdisplay: An Analysis and Reconstruction Package for VERITAS

DOI 10.5281/zenodo.14283615 License BSD 3-Clause CI passing

- DL1 from data (stored at UCLA and DESY): ~50 TB *internal*
- DL1 from simulations (stored at DESY): ~250 TB *internal*

VERITAS DL2 - Showers

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)

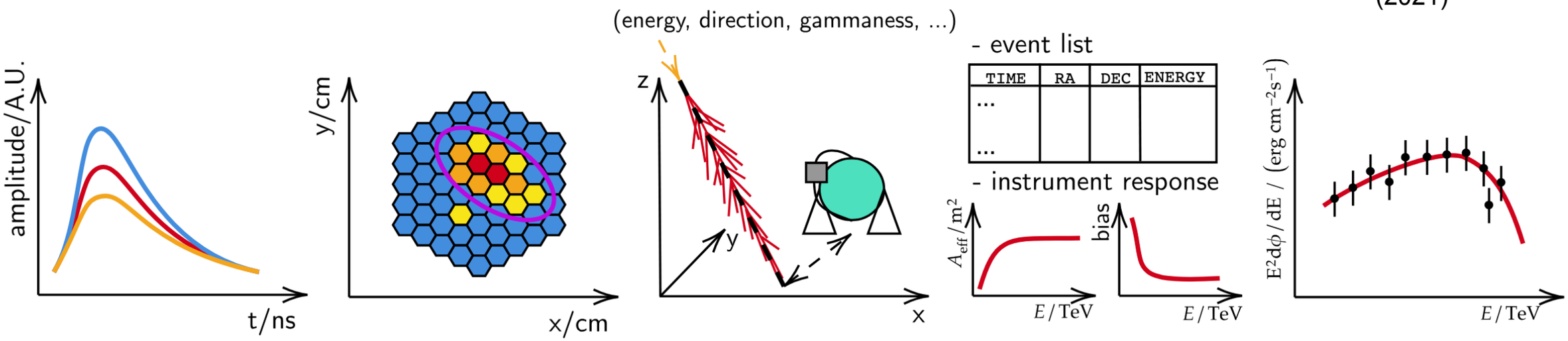


- Shower parameters (ROOT-format)

- DL2 from data (stored at UCLA and DESY): 10 TB *internal*
- DL2 from simulations: not stored *transient*

VERITAS DL3 - Events and response functions

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)



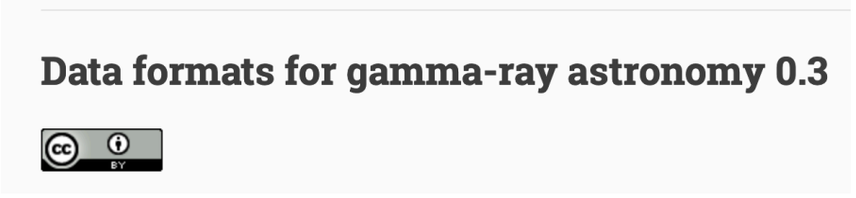
DL0 raw DAQ output → **DL1** image parameters → **DL2** shower parameters → **DL3** event list + IRF → **DL4** science products

- Community standardized data format implemented as FITS. *public*
  - Event lists, instrument response functions
- Community developed science software (gammapy); software for CTAO. *public*
- Converters from internal VERITAS format to DL3-FITS. *public*

- DL3 for data: ~ 4GB per gamma/hadron cut (5-7 cuts). *internal*
- “Internal publication” in private Zenodo repository in progress. *almost (!) public*

Data formats for gamma-ray astronomy GDAF

- Community developed data format
  - Close to NASA / OGIP / HEASARC FITS standards
- Per observation (run)
  - Tables
    - Event list with energy, direction, time, gamma/hadron separator
    - Instrument response functions (effective areas, energy / angular migration matrix, background)
    - Good time intervals
  - Header
    - Data quality information (e.g., weather, rates)
    - Metadata (including pointing)



250 kB per observation including all necessary information for a gamma-ray analysis

VERITAS DL3 - Metadata

```
HDUCLASS= 'GADF    '           / This FITS file follows the GADF data format
HDUDOC  = 'https://gamma-astro-data-formats.readthedocs.io'
HDUVERS = '0.3     '           / DL3 specification version
HDUCLAS1= 'EVENTS  '           / Primary extension class
RADECSYS= 'FK5     '           / equatorial system type
EQUINOX =                2000.0 / base equinox
CREATOR = 'pyV2DL3 v0.5::EventDisplay'
ORIGIN  = 'VERITAS Collaboration' / Data from VERITAS
TELESCOP= 'VERITAS '
INSTRUME= 'Epoch V6_2012_2013a'
```

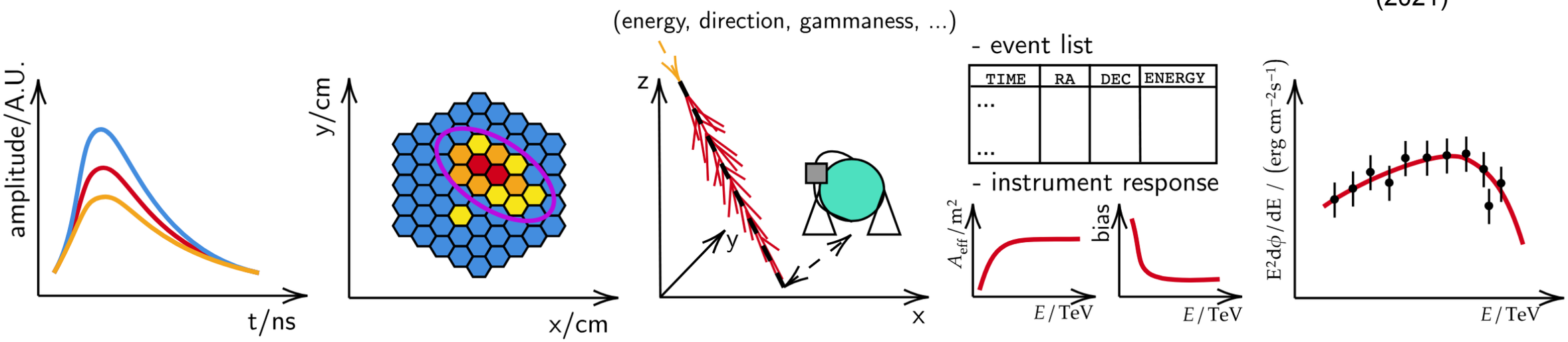
```
OBS_ID  =                 64080 / Run Number
DATE-OBS= '2012-10-13T10:52:08.109' / start date (UT
DATE-AVG= '2012-10-13T11:02:09.002' / average date (
DATE-END= '2012-10-13T11:12:09.895' / end date (UTC)
TSTART  =    242909531.1089129 / mission time of sta
TSTOP   =    242910732.8951768 / mission time of end
MJDREFI =              53402.0 / int part of referen
```

```
WEATHER = 'A       '           / weather conditions (A is best)
CONFIG  =                   15 / telescope configuration mask
TRIGCFG = 'normal  '           / trigger configuration
DATACAT = 'science '           / data category (e.g., science, calibr
DQMSTAT = 'good_run'           / DQM status (e.g., good, do_no_use)
DQMREAS = 'NULL    '           / DQM status reason
DQMMASK = 'NULL    '           / DQM telescope cut mask
VPMCFG  =                   15 / VPM configuration mask
LIGHTLEV= 'dark    '           / light level (from currents)
L3RATE  =   444.6298783182443 / mean L3 rate [Hz]
L3RATESD=   20.87659627614281 / std deviation of L3 rate [Hz]
CURRMEAN=   8.301287758015462 / mean currents [muA]
CURRSTD =   3.562403206978189 / std deviation of currents [muA]
CURRMED =                 8.0 / median currents [muA]
WINDSPE =   5.977619047619047 / mean wind speed [mph]
WINDMAX =   7.114428571428571 / maximum wind speed [mph]
WINDMIN =   4.738857142857143 / minimum wind speed [mph]
WINDDIR =   84.00761904761904 / wind direction [deg]
AIRTEMP =   48.63095238095238 / air temperature [deg_C]
RELHUMID=   61.75428571428572 / relative humidity [pct]
FIRMEAN0=  -31.29761904761904 / mean FIR temperature (TEL0) [deg_C]
FIRMEAN1=  -35.46111111111111 / mean FIR temperature (TEL1) [deg_C]
```

Try to be as generous as possible
In terms of what data required for data quality selection might be required

VERITAS Catalogue VTSCat - DL4

Nigro, C., Tarek, H, Olivera-Nieta, L. (2021)



DL0 raw DAQ output → **DL1** image parameters → **DL2** shower parameters → **DL3** event list + IRF → **DL4** science products

- Science products consist of spectra, light curves, sky maps, …
- every single result of VERITAS is available in machine readable format
- VERITAS Catalogue **VTSCat.** *public*
- Simple human-readable text files in (e)CSV and YAML
- (Not a catalogue)

Paper: https://arxiv.org/pdf/2301.04498
HEASARC: https://heasarc.gsfc.nasa.gov/W3Browse/veritas/verimaster.html
Zenodo: https://zenodo.org/records/8386703
GitHUB: https://github.com/VERITAS-Observatory/VERITAS-VTSCat

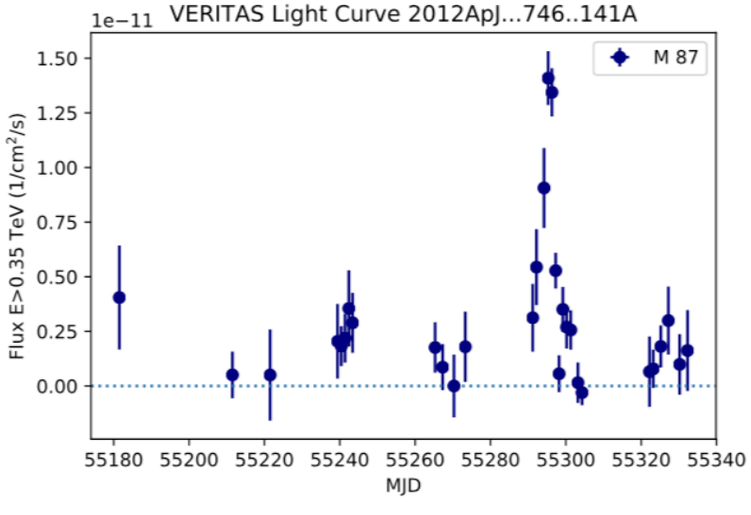
VTSCat Example

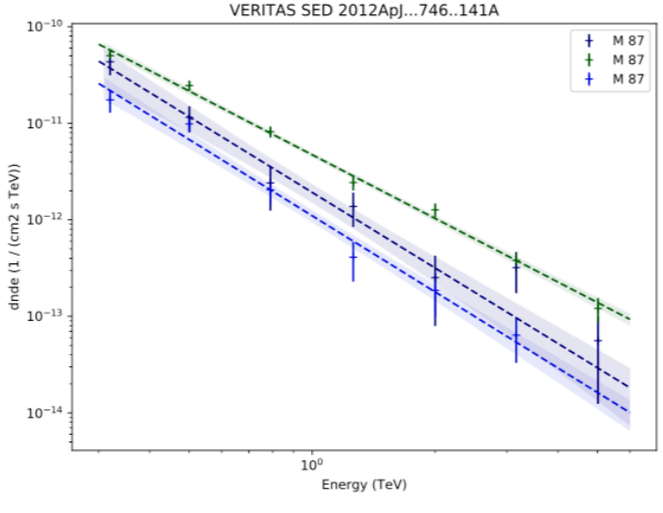
2012ApJ...746..141A: VERITAS Observations of Day-scale Flaring of M 87 in 2010 April

```
# %ECSV 0.9
# ---
# datatype:
# - {name: time, unit: MJD, datatype: float64}
# - {name: livetime, unit: day, datatype: float64}
# - {name: e_min, unit: TeV, datatype: float64}
# - {name: flux, unit: cm-2 s-1, datatype: float64}
# - {name: flux_err, unit: cm-2 s-1, datatype: float64}
# meta: !!omap
# - data_type: lc
# - source_id: 58
# - reference_id: 2012ApJ...746..141A
# - telescope: veritas
# - {SED_TYPE: flux}
time livetime e_min flux flux_err
55181.5    0.00900000  0.35   4.04900e-12    2.38100e-12
55211.5     0.0140000  0.35   5.07600e-13    1.06800e-12
55221.5     0.0140000  0.35   5.02400e-13    2.08700e-12
55239.4     0.0450000  0.35   2.04400e-12    1.70900e-1
55240.4     0.0510000  0.35   1.82300e-12    9.12200e-1
55241.4     0.0220000  0.35   2.19000e-12    1.11900e-1
55242.4     0.0290000  0.35   3.54900e-12    1.74500e-1
55243.4     0.0150000  0.35   2.89100e-12    1.37300e-1
55265.3     0.0150000  0.35   1.76200e-12    1.15500e-1
55267.3     0.0210000  0.35   8.61700e-13    1.05300e-1
55270.3     0.0210000  0.35       0.00000    1.43600e-1
55273.3     0.0140000  0.35   1.79400e-12    1.61000e-1
55291.2     0.0210000  0.35   3.12200e-12    1.55200e-1
55292.2     0.0150000  0.35   5.43800e-12    1.73900e-1
55294.2     0.0280000  0.35   9.06300e-12    1.83200e-12
55295.3      0.104000  0.35   1.40900e-11    1.23400e-12
```

```
# %ECSV 0.9
# ---
# datatype:
# - {name: e_ref, unit: TeV, datatype: float32}
# - {name: dnde, unit: m-2 s-1 TeV-1, datatype:
# - {name: dnde_err, unit: m-2 s-1 TeV-1, datat
# meta: !!omap
# - data_type: sed
# - source_id: 58
# - file_id: 1
# - reference_id: 2012ApJ...746..141A
# - telescope: veritas
# - mjd: {max: 55294.0, min: 55291.0}
# - comments: |
#     Figure 4 (rising)
e_ref dnde dnde_err
0.32    4.3294e-07    1.1965e-07
0.50    1.1513e-07    3.4768e-08
0.79    2.4097e-08    1.1666e-08
1.26    1.3761e-08    5.3340e-09
2.00    2.5202e-09    1.7264e-09
3.16    3.1948e-09    1.4541e-09
5.01    5.5951e-10    4.3542e-10
```

Name
- ..
- figures
- README.md
- VER-000058-1.yaml
- VER-000058-2.yaml
- VER-000058-3.yaml
- VER-000058-lc-1.ecsv
- VER-000058-lc-2.ecsv
- VER-000058-lc-3.ecsv
- VER-000058-lc-4.ecsv
- VER-000058-lc-5.ecsv
- VER-000058-lc-6.ecsv
- VER-000058-lc-7.ecsv
- VER-000058-sed-1.ecsv
- VER-000058-sed-2.ecsv
- VER-000058-sed-3.ecsv
- info.yaml





Format inspired by gamma cat

Conclusions - VERITAS Data Publication

- Notable progress in publication of data and software.
  - Prepare most data/software for data publication (even without decision on publication)

- Long-term software and data preservation challenging for data from “old” collaborations.
  - Drain of expertise.
  - Cases of unclear licenses.
  - Data not free of systematics and expert knowledge hard to replace.

- CTAO is a driver for the development of community standards.
  - Simplification of data analysis for newcomers.