Astronomy & Astrophysics manuscript no. output
August 9, 2019

Towards open and reproducible multi-instrument analysis in gamma-ray astronomy

C. Nigro[1⋆], C. Deil[2], R. Zanin[2], T. Hassan[1], J. King[3], J.E. Ruiz[4], L. Saha[5], R. Terrier[6], K. Brügge[7], M. Nöthe[7], R. Bird[8], T. T. Y. Lin[9], J. Aleksić[10], C. Boisson[11], J.L. Contreras[5], A. Donath[2], L. Jouvin[10], N. Kelley-Hoskins[1], B. Khelifi[6], K. Kosack[12], J. Rico[10], and A. Sinha[6]

(Affiliations can be found after the references)

August 9, 2019

ABSTRACT

The analysis and combination of data from different gamma-ray instruments involves the use of collaboration proprietary software and case-by-case methods. The effort of defining a common data format for high-level data, namely event lists and instrument response functions (IRFs), has recently started for very-high-energy gamma-ray instruments, driven by the upcoming Cherenkov Telescope Array (CTA). In this work we implemented this prototypical data format for a small set of MAGIC, VERITAS, FACT, and H.E.S.S. Crab nebula observations, and we analyzed them with the open-source `gammapy` software package. By combining data from *Fermi*-LAT, and from four of the currently operating imaging atmospheric Cherenkov telescopes, we produced a joint maximum likelihood fit of the Crab nebula spectrum. Aspects of the statistical errors and the evaluation of systematic uncertainty are also commented upon, along with the release format of spectral measurements. The results presented in this work are obtained using open-access on-line assets that allow for a long-term reproducibility of the results.

Key words. Methods: data analysis, Gamma rays: general

1. Introduction

The opening of new astronomical windows at different wavelengths in the last decades has made evident that many astrophysical puzzles could be solved only by combining images obtained by different facilities. In the late 1970s a common format was developed to facilitate the image interchange between observatories, hence overcoming incompatibilities between the numerous operating systems. The Flexible Image Transport System (FITS) format was standardized in 1980 (Wells et al. 1981) and formally endorsed in 1982 by the International Astronomical Union (IAU), that in 1988 formed a FITS working group (IAUFWG) entrusted to maintain the format and review future extensions. In the mid-1990s the NASA high energy astrophysics science archive research center (HEASARC) FITS Working Group, also known as the OGIP (Office of Guest Investigator Programs), promoted multi-mission standards for the format of FITS data files in high-energy astrophysics and produced a number of documents and recommendations that were subsequently incorporated into the FITS standard format definition. Since its conception, the FITS format has been been updated regularly to address new types of metadata conventions, the diversity of research projects and data product types. The last version of the FITS Standard document (4.0) was released in 2018[1].

Today the FITS format is in widespread use among astronomers of all observing bands, from radio frequencies to gamma rays. For instance, the high-energy gamma-ray (HE, E>100 MeV) Large Area Telescope (LAT, Atwood et al. 2009), on board of the *Fermi* satellite, publicly releases all its high-level analysis data in FITS format, that, processed with the science tools, are used to obtain scientific products as spectra, light-curves and sky-maps. However, as a branch of astroparticle physics, very-high-energy (VHE, E>100 GeV) gamma-ray astronomy inherited its methodologies and standards from particle physics, where the ROOT[2] framework (Brun & Rademakers 1997) and its associated file format is commonly used. Despite the common container format neither the internal data structure nor the software is shared among the different experiments. VHE gamma-ray astronomy is conducted by ground-based telescopes, with the Imaging atmospheric Cherenkov telescopes (IACTs) among the most successful ones (de Naurois & Mazin 2015). Data from four of the currently operating IACTs were used in this project: the Major Atmospheric Gamma Imaging Cherenkov Telescopes (MAGIC, Aleksić et al. 2016), the Very Energetic Radiation Imaging Telescope Array System (VERITAS, Holder et al. 2006), the First G-APD Cherenkov Telescope (FACT, Anderhub et al. 2013) and the High Energy Stereoscopic System (H.E.S.S., Hinton et al. 2004). Each of them is described in section 2.

A new era in VHE gamma-ray astronomy is expected to start with the future Cherenkov Telescope Array (CTA, Acharya et al. 2013), the next generation IACT instrument, which is currently under construction. The future operation of CTA as an observatory calls for its data format and analysis software to be available to a wide astronomical community. This requirement led to a standardization of the IACT data format, adopting the FITS standard, and the development of open-source science tools, initiating the integration of the VHE discipline into multi-instrument astronomy. A first attempt to define a common data format for the VHE gamma-ray data is being carried out within the "Data for-

⋆ *Send offprint requests to*: C. Nigro, `cosimo.nigro@desy.de`

[1] `https://fits.gsfc.nasa.gov/fits_standard.html`

[2] `https://root.cern.ch/`

mats for gamma-ray astronomy"[3] forum (Deil et al. 2017b, 2018). This is a community effort clustering together members of different IACT collaborations with CTA as driving force. In this paper we implement this prototypical data format for data samples by MAGIC, VERITAS, FACT, and H.E.S.S. and we combine, for the first time, observations by *Fermi*-LAT and these four IACTs relying for the scientific analysis solely on open-source software, in particular on the `gammapy`[4] project (Donath et al. 2015; Deil et al. 2017a). We provide the reader not only with the datasets but also with all the scripts and an interactive environment to reproduce all the results. They are available at `https://github.com/open-gamma-ray-astro/joint-crab` and will be referred to, from now on, as online material. This allows to fully reproduce the results presented in the paper. The Crab nebula is selected as a target source for this work, being the reference source in the VHE gamma-ray astronomy (Aharonian et al. 2004, 2006; Albert et al. 2008; Aleksić et al. 2012; Aleksić et al. 2016) due to its brightness, apparent flux steadiness and visibility from all the considered observatories.

The paper is structured as follows: we describe the selected datasets in Section 2, the process of extracting the spectral information with some considerations on the handling of statistical and systematic uncertainties in Section 3. In Section 4 we present the resources we use to ensure the analysis reproducibility, and in Section 5 we provide some prospects for the reuse of the methodologies of data release discussed and implemented in this work. This report is a technical paper to show the ease of multi-instrument results once the format standardization is reached: we do not seek to draw any scientific conclusion on the physics of the Crab nebula, and of pulsar wind nebulae, in general.

2. Datasets

Differently than other astronomical telescopes, instruments for gamma astronomy cannot directly scatter or reflect gamma rays, being photon-matter interactions dominated by pair production for $E_\gamma > 30\,\mathrm{MeV}$. The experimental techniques, either space-borne or ground-based (Funk 2015), rely on the direct detection of secondary charged particles or on the indirect detection of the Cherenkov emission of a cascade of charged secondaries they produce in the atmosphere. A detection, or event, cannot be unambiguously discriminated from the irreducible charged cosmic ray background, but can only be classified with a certain probability as a primary photon. Gamma-ray astronomy could therefore be labelled as "event-based" in contrast to the "image-based" astronomy where charge-coupled devices (CCDs) act as detectors. The input for the high-level analysis of gamma-ray astronomy data is typically constituted by two elements. The first one is a list of events that are classified (according to selection cuts) as gamma rays, along with their estimated direction, $\boldsymbol{P'}$, estimated energy, E', and arrival time. The second element consists of the instrument response functions (IRFs), quantifying the performance of the detector and connecting estimated quantities (E', $\boldsymbol{P'}$) with their true, physical, values (E, $\boldsymbol{P}$). IRFs components are:

- effective area, representing the effective collection area of the instrument, $A_{\mathrm{eff}}(E, \boldsymbol{P})$;
- energy dispersion, the probability density function of the energy estimator $f_E(E'|E, \boldsymbol{P})$;
- point spread function (PSF), the spatial probability distribution of the estimated event directions for a point source, $f_P(\boldsymbol{P'}|E, \boldsymbol{P})$.

Their formal definition is shared with lower-energy instruments (e.g. x-ray, Davis 2001) and they are computed from Monte Carlo simulations. Since the detector response is not uniform across the field of view (FoV), the IRFs generally depend on the radial offset from the FoV center (full-enclosure IRFs). When this dependency is not taken into account, and a cut on the direction offset of the simulated events is applied, the IRFs are suited only for the analysis of a point-like source sitting at an a priori defined position in the FoV (point-like IRFs). IRFs components, in this case, do not have a dependency on the event coordinate, $\boldsymbol{P}$, but only on the energy and the PSF component is not specified. The differences between full-enclosure and point-like IRFs are illustrated in the interactive notebook `1_data.ipynb` in the online material. For this publication we will use all the datasets to perform a point-like analysis. In the IACT terminology, event lists and IRFs are dubbed Data Level 3 (DL3) products (Contreras et al. 2015). The datasets used in this work include observations of the Crab nebula by *Fermi*-LAT, MAGIC, VERITAS, FACT, and H.E.S.S. following the format specifications available in the "Data formats for gamma-ray astronomy" forum (Deil et al. 2017b). The IACT DL3 datasets were produced with proprietary codes that extracted the event lists, and IRFs, and saved them in the requested format[5]. They are released in chunks, typically of 20-30 minutes, of data acquisition, named runs. IACTs are ground-based, pointing instruments and their response varies with the observing conditions (atmospheric transmission, zenith angle, night sky background level, position of the source in the FoV) therefore their data come with per-run IRFs. The *Fermi*-LAT telescope, orbiting around the Earth at $\sim 600\,\mathrm{km}$, is generally operating in survey mode and has a stable set of IRFs shipped with the science tools. We use the *Fermi*-LAT science tools and `gammapy` to make these datasets spec-compliant. This work constitutes the first joint release of datasets from different instruments in VHE gamma-ray astronomy.

An interactive notebook illustrating the content of the datasets, `1_data.ipynb`, is available in the online material. The datasets are presented in what follows in order of increasing instrument energy threshold (see Table 1).

Table 1. Crab nebula datasets used in this work. T_{obs} stands for observation time, E_{min} and E_{max} identify the energy range of the analysis, N_{on} and N_{bkg} the number of total and background events, respectively, estimated in the circular signal region with a radius R_{on}.

Dataset	T_{obs}	E_{min} TeV	E_{max} TeV	N_{on}	N_{bkg}	R_{on} deg
Fermi-LAT	~7 yr	0.03	2	578	1.2	0.30
MAGIC	0.66 h	0.08	30	784	129.9	0.14
VERITAS	0.67 h	0.16	30	289	13.7	0.10
FACT	10.33 h	0.45	30	691	272.8	0.17
H.E.S.S.	1.87 h	0.71	30	459	27.5	0.11

[3] `https://gamma-astro-data-formats.readthedocs.io`
[4] `https://gammapy.org/`
[5] All the FACT software used to generate DL3 datasets is open-source.

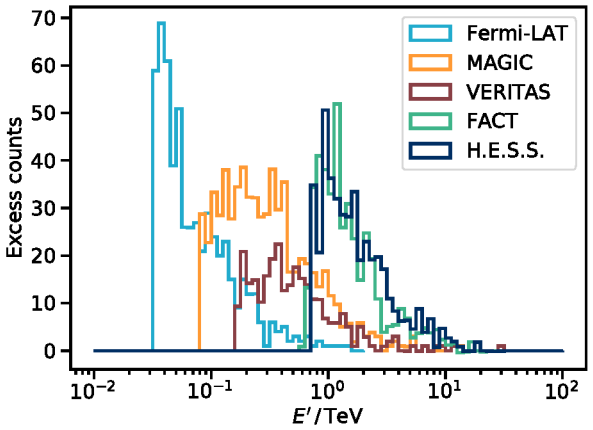

Fig. 1. Histograms of the estimated mean number of signal events from the Crab nebula, number of excess events, vs estimated energy for each dataset.

2.1. Fermi-LAT

The LAT detector, on board of the *Fermi* satellite, is an imaging pair-conversion telescope, which has been designed to detect photons between 20 MeV and more than 300 GeV (Atwood et al. 2009). We analyzed the publicly available observations spanning from 2008 August 8 to 2015 August 2, i.e. ∼7 years of operations, in a 30°-radius region around the position of the Crab nebula. We used all FRONT and BACK type events belonging to the Source class with a reconstructed direction within 105° from the local zenith (to reject the emission from the Earth's atmosphere) and a reconstructed energy between 30 GeV and 2 TeV. The lower energy cut was chosen to minimize the contamination of the Crab pulsar that is located at the center of the nebula. We estimated that above 30 GeV the pulsar emission contributes to less than 10% to the detected flux and can be neglected given the technical purpose of the paper. An interactive notebook illustrating this calculation is available in the online material (`5_crab_pulsar_nebula_sed.ipynb`). To reduce the IRFs to a DL3-compliant format, we compute the PSF using `gtpsf` and estimated the effective area from the sky-coordinate and energy-dependent exposure computed with `gtexpcube2`. The effective area is simply the exposure scaled by the observation time. The energy dispersion at the energies considered in this analysis has an impact smaller than 5% on the reconstructed spectrum[6]. We approximate the energy dispersion with a Gaussian distribution with mean (bias) 0 and standard deviation 0.05 (resolution) independent on the estimated direction and energy. Event list and IRFs produced by the *Fermi*-LAT science tools are already FITS files, `gammapy` is used to store them in a DL3-compliant format.

2.2. MAGIC

MAGIC is a system of two 17-m diameter IACTs, with 3.5° FoV located on the Canary island of La Palma, Spain at the Roque de Los Muchachos Observatory (2200 m above sea level). The first telescope worked in stand-alone mode from 2004 until 2009, when a second one started operations. In 2012 the two MAGIC telescopes underwent a major upgrade of the readout systems and the camera of the first telescope (Aleksić et al. 2016) leading to a significant improvement of the instrument performance (Aleksić et al. 2016). The FITS data released by the MAGIC collaboration includes 40 min of Crab nebula observations and their corresponding IRFs. The data were recorded in 2013, after the aforementioned major upgrade, at small zenith angles (< 30°) in wobble mode (Fomin et al. 1994) with the source sitting at an offset of 0.4° from the FoV center. They are chosen from the data sample used for the performance evaluation in Aleksić et al. (2016). The IRFs released by the MAGIC collaboration were generated using the MARS software (Zanin et al. 2013), for a point-like source response: i.e. they are calculated for a simulated source at 0.4° offset, applying a directional cut on the events direction of 0.14° around the source position.

2.3. VERITAS

VERITAS consists of four 12-m diameter IACTs (Holder et al. 2006), with 3.5° FoV based at the Fred Lawrence Whipple Observatory in Southern Arizona, USA. Since the start of full array operations in 2007, the sensitivity of VERITAS was improved by two major upgrades: the relocation of one of its telescopes in 2009 (Perkins et al. 2009) and the upgrade of the camera with higher quantum efficiency photomultiplier tubes in 2012 (Kieda et al. 2013). The DL3 data released by the VERITAS collaboration includes 40 min of archival observations of the Crab nebula taken in 2011, after the telescope relocation, but prior to the camera upgrade. These observations were carried out in wobble mode with the standard offset of 0.5° at small zenith angles (< 20°). The released IRFs are valid for the analysis of point-like sources taken with the standard offset angle and were generated from events surviving the standard directional cut of 0.1°, using the VEGAS software package (Cogan et al. 2008).

2.4. FACT

FACT (Anderhub et al. 2013; Biland et al. 2014) is a single IACT with 4-m diameter reflective surface and a FoV of 4.5°, that is located next to MAGIC at the Roque de Los Muchachos Observatory. FACT tests the feasibility of silicon photomultipliers (SiPM) for use in VHE gamma-ray astronomy. It is the first fully automated Cherenkov telescope that takes data without an operator on site (Nöthe et al. 2018). For this work FACT released one full week of observations of the Crab nebula taken in November 2013, corresponding to 10.3 h[7]. The data were recorded in wobble mode with an offset angle of 0.6° at zenith angles smaller than 30°. The corresponding IRFs are of point-like type with a directional cut of 0.17°.

2.5. H.E.S.S.

H.E.S.S. is an array of five IACTs located in Namibia, on the Khomas Highland, near the Gamsberg mountain. The first four 12-m diameter telescopes, arranged in a square, became operational in December 2003 marking the start of what today is called H.E.S.S. Phase I with a FoV of 5°. Since July 2012 a fifth 28-m diameter telescope, located at the array center, started operation (H.E.S.S. Phase II) both in stereoscopic and stand-alone mode. In this work, we used four observation runs of the Crab nebula carried out by H.E.S.S. Phase I in 2004,

[6] `https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Pass8_edisp_usage.html`

[7] `https://factdata.app.tu-dortmund.de/`

each of them with a duration of 28 min. They were taken in wobble mode at zenith angles between 45° and 50°, half of them with a 0.5°, and the other half with a 1.5° offset angle. These data are the Crab runs part of the first FITS test data release (Abdalla et al. 2018)[8]. H.E.S.S. released full-enclosure IRFs (Deil et al. 2017b), i.e. no directional cut is applied on the simulated events.

3. Data analysis

In this section we present a spectral analysis of the gamma-ray datasets described in Section 2. First, the gamma-ray event data and IRFs are reduced for each instrument (Section 3.1). Then, in Section 3.2, we preform a spectral likelihood fit, under the assumption of a log-parabola analytic model, for each datasets separately, and for all the datasets together (joint fit). Finally, we present an analysis that includes a systematic error term, representing the uncertainty on the energy scale of each instrument, in a modified likelihood function.

3.1. Spectrum extraction

In order to estimate the energy spectrum of a gamma-ray source ($\frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\mathbf{\Lambda})$, with $\mathbf{\Lambda}$ a the set of spectral parameters), a binned maximum likelihood method, with $n_{E'}$ bins in estimated energy E' is used. The observed data $\mathbf{D}$ for such a likelihood function are the number of events in a circular signal region (labeled as ON) containing the gamma-ray source and in a control region (labeled as OFF) measuring the background to be subtracted from the ON. Considering n_{runs} observation runs from n_{instr} different instruments (or datasets), we can write the likelihood as:

$$\mathcal{L}(\mathbf{\Lambda}|\mathbf{D}) = \prod_{i=1}^{n_{\mathrm{instr}}} \mathcal{L}_i(\mathbf{\Lambda}|\{N_{\mathrm{on},ijk}, N_{\mathrm{off},ijk}\}_{j=1,\dots,n_{\mathrm{runs}};k=1,\dots,n_{E'}}) \quad (1)$$

with each instrument contributing with a term:

$$\mathcal{L}_i(\mathbf{\Lambda}|\{N_{\mathrm{on},ijk}, N_{\mathrm{off},ijk}\}_{j=1,\dots,n_{\mathrm{runs}};k=1,\dots,n_{E'}}) = \prod_{j=1}^{n_{\mathrm{runs}}}\prod_{k=1}^{n_{E'}} \mathrm{Pois}(g_{ijk}(\mathbf{\Lambda}) + b_{ijk}; N_{\mathrm{on},ijk}) \times \mathrm{Pois}(b_{ijk}/\alpha_{ij}; N_{\mathrm{off},ijk}), \quad (2)$$

where:

- $N_{\mathrm{on},ijk}$ and $N_{\mathrm{off},ijk}$ are the number of observed events within the ON and OFF regions, respectively, in the energy bin k in the run j for the i-th instrument. They are both characterized by a Poisson distribution;
- $g_{ijk}(\mathbf{\Lambda})$ and b_{ijk} are the expected number of signal and background events, respectively, in the energy bin k in the run j for the i-th instrument. g_{ijk} is computed with the forward folding technique: for a point-like analysis the assumed spectrum $\frac{\mathrm{d}\phi}{\mathrm{d}E}$ is convolved with the effective area and energy dispersion IRFs component. b_{ijk}, in absence of a background spectral model, is treated as a nuisance parameter and fixed to the value returning $\frac{\partial\mathcal{L}}{\partial b_{ijk}} = 0$, for the mathematical details of g_{ijk} and b_{ijk} evaluation see Appendix A in Piron et al. (2001).
- α_{ij} is the ON to OFF exposures ratio, constant with energy in our case, in the run j for the i-th instrument.

The size of the circular ON region per each dataset is given as R_{on} in Table 1, the OFF region can be defined either as multiple regions mirroring the ON symmetrically with respect to the telescope pointing position (reflected regions background Aharonian et al. 2001; Berge, D. et al. 2007) or as a ring around the source position (ring background method in Berge, D. et al. 2007). Given the wobble mode observation strategy, and the small FoV, the reflected regions method is naturally suitable for the IACTs datasets. On the other hand, the ring background method is used for the *Fermi*-LAT datasets, in this analysis with a circular signal region of 0.3° radius and a background ring with inner and outer radius of 1° and 2°, respectively. We choose, for all the instruments, 20 bins per decade for the estimated energy between 10 GeV and 100 TeV. For a given instrument i and run j the likelihood values are not computed in the energy bins outside the range $[E_{\mathrm{min}}, E_{\mathrm{max}}]$ given in Table 1. The choice of the energy range for *Fermi*-LAT is already discussed in Section 2.1. For the IACT datasets, E_{min} is a safe energy threshold for the spectrum extraction computed by the collaboration software and hard coded in the DL3 files. It is mostly dependent on the experiment performance and on the zenith angle of the observations. FACT has an energy threshold of 450 GeV, higher than MAGIC and VERITAS despite the observations carried out in the same zenith angle range, due to its limited light-collection area and the single telescope observations. The larger zenith angle of the H.E.S.S. datasets is due to the low altitude at which the source culminates at the H.E.S.S. site. This yields an energy threshold of $\sim$ 700 GeV, higher than any other IACT. The maximum energy E_{max} is fixed to 30 TeV for all the IACTs and it is chosen to cover the whole energy range containing events.
The mean number of signal events, labeled as excess events, can be estimated via the equation $N_{\mathrm{ex},ijk} = N_{\mathrm{on},ijk} - \alpha_{ij}N_{\mathrm{off},ijk}$. The distribution of the excess events in each estimated energy bin, summed over all the observational runs per each instrument ($\sum_{j=1}^{n_{\mathrm{runs}}} N_{\mathrm{ex},ijk}$) is shown in Figure 1. Table 1 reports also the total number of observed events in the ON region ($N_{\mathrm{ON}} = \sum_{ijk} N_{\mathrm{on},ijk}$) per each experiment, and the number of background events in the ON region per each experiment, obtained scaling the OFF events with the exposures ratio α_{ij} ($N_{\mathrm{bkg}} = \sum_{ijk} \alpha_{ij} N_{\mathrm{off},ijk}$).
To perform a joint point-like analysis we reduce the *Fermi*-LAT and H.E.S.S. full-enclosure IRFs to a point-like format, removing the dependency from the source position in the FoV. For *Fermi*-LAT, under the assumption that the acceptance is uniform in a small sky region close to our target, we obtained a point-like effective area by taking the value at the source position. In each energy bin we corrected the effective area with a containment fraction computed integrating the PSF over the signal region. Similarly for the H.E.S.S. IRFs the value at the source offset is taken and a correction based on the PSF containment fraction is computed.

3.2. Spectral model fit

We assumed a spectral model of log-parabolic form:

$$\frac{\mathrm{d}\phi}{\mathrm{d}E} = \phi_0 \left(\frac{E}{E_0}\right)^{-\Gamma-\beta\log_{10}\left(\frac{E}{E_0}\right)}, \quad (3)$$

since it was suggested as the best-function approximation for the Crab nebula spectrum in such a wide energy range (Alek-

[8] https://www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1/

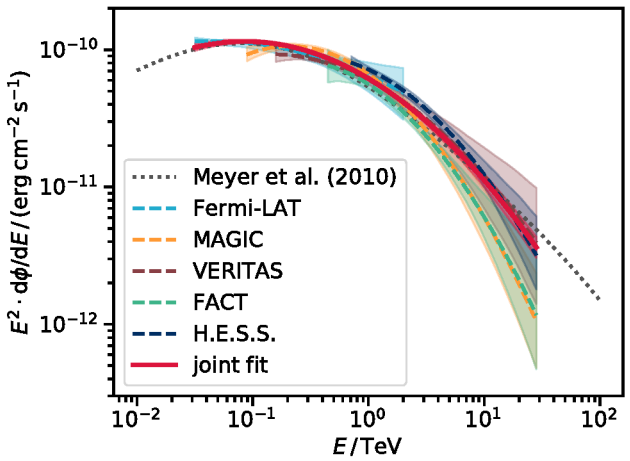

Fig. 2. Crab nebula SED for individual instrument fits and from the joint fit. Single-instrument results are represented with dashed lines, the fit of all the datasets together, labeled as joint, is represented as a thick, solid red line. The shaded areas represent the SED error bands whose calculation is explained in Section 3.2. The dotted line shows the model in Meyer et al. (2010).

sić et al. 2015). The spectral parameters $\mathbf{\Lambda} = (\phi_0, \Gamma, \beta)$ are left free to vary in the fit while the reference energy E_0 is fixed at the value of 1 TeV. We refer to the result of the maximum likelihood using all the instrument datasets as *joint fit*. As a consistency check we also fitted each instrument dataset separately (i fixed in Eq. 1). E_0 is usually chosen to minimize the correlation between the other spectral parameters. In this work, in order to directly compare the parameters $\mathbf{\Lambda}$ also for the fit with the individual instrument datasets, the reference energy E_0 is kept fixed at the same value of 1 TeV. This introduces larger errors and correlation for the datasets for which such value is close to one of the extremes of its energy range. The resulting spectral energy distributions (SEDs, $E^2 \mathrm{d}\phi/\mathrm{d}E$) are shown in Fig. 2, together with a theoretical model taken from Meyer et al. (2010). The values of the fit parameters are listed in Table 2. The joint fit inherently comes with an increase in statistical power, as evidenced by the shrinking of the confidence contours of the fitted spectral parameters for the joint fit in Fig. 3. We note that the Crab SED shape is not exactly represented by log parabola across the 30 GeV to 20 TeV energy range, which is one reason for differences in the measured fit parameters from the different experiments. An interactive summary of the spectral results is available in the online material (`2_results.ipynb`).

Table 2. Spectral model best-fit parameters, as defined in Eq. 3. The reference energy, E_0, is taken at 1 TeV for all the fit results. The prefactor ϕ_0 is given in 10^{-11} TeV^{-1} cm^{-2} s^{-1}.

Dataset	ϕ_0	Γ	β
Fermi-LAT	4.04 ± 1.01	2.37 ± 0.24	0.14 ± 0.13
MAGIC	4.15 ± 0.30	2.60 ± 0.10	0.44 ± 0.11
VERITAS	3.76 ± 0.36	2.44 ± 0.09	0.26 ± 0.17
FACT	3.49 ± 0.30	2.54 ± 0.22	0.42 ± 0.31
H.E.S.S.	4.47 ± 0.29	2.39 ± 0.18	0.37 ± 0.22
joint	3.85 ± 0.11	2.51 ± 0.03	0.24 ± 0.02

The statistical uncertainty on the SED is estimated by using a sampling technique to propagate the errors from the fit parameters. We assume that the likelihood of the model parameters $\mathbf{\Lambda}$ is distributed according to a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ defined by the fit results. We assume $\boldsymbol{\mu} = \hat{\mathbf{\Lambda}} = (\hat{\phi_0}, \hat{\Gamma}, \hat{\beta})$, values of the fitted parameters and $\mathbf{\Sigma} = \hat{\mathbf{V}}_{\hat{\mathbf{\Lambda}}}$, covariance of the fitted parameters. We sample this distribution and compute the spectrum realization corresponding to each sampled $\mathbf{\Lambda}$. The $\pm 1\,\sigma$ uncertainty on the fitted spectrum is estimated by taking, at a given energy, the quantiles of the fluxes distribution that returns a 68% containment of all the realizations. The upper and lower limits of the error band estimated with our method are plotted in black against 100 realizations of the spectrum with sampled parameters in Fig. 4, for the example case of the VERITAS datasets.

3.3. Systematic uncertainties on different energy scales

Spectral measurements in gamma-ray astronomy are affected by multiple sources of systematic uncertainty. The DL3 data contains systematic uncertainties that originate from an imperfect modeling of the atmosphere, telescopes and event reconstruction, resulting in a shift of the reconstructed energy scale and errors in the assumed IRF shapes. A second source of systematic error comes from the data reduction and model fitting, e.g. due to energy binning and interpolation effects, as well as from source morphology and spectral shape assumptions. Generally two approaches are used to evaluate and report systematic errors (Conrad et al. 2003; Barlow 2017): multiple analyses / bracketing as in (Aharonian et al. 2006; Aleksić et al. 2012) or modified likelihood with nuisance parameters (Dickinson & Conrad 2013; Dembinski et al. 2017; Ballet et al. 2018). The first approach leads to the estimation of an overall systematic error on each spectral parameter, e.g. for the flux normalization $\phi_0 \pm \sigma_{\phi_0,\mathrm{stat.}} \pm \sigma_{\phi_0,\mathrm{syst.}}$, whereas the second method yields a global error including both statistical and systematic uncertainties, i.e. $\phi_0 \pm \sigma_{\phi_0,\mathrm{stat.+syst.}}$. As an example of how to treat systematic errors, we present here an analysis with a modified likelihood that includes the uncertainty on the energy scale. Following Dembinski et al. (2017), we define a new joint likelihood function that includes a constant relative bias of the energy estimator per each instrument z_i, characterized by a Gaussian distribution with mean 0 and standard deviation δ_i, the systematic uncertainty on the energy scale estimated by the single instrument ($\mathcal{N}(z_i; 0, \delta_i^2)$ in the following notation). This parameter is defined as $z_i = \frac{\tilde{E}-E}{E} = \frac{\tilde{E}}{E} - 1$, with $\tilde{E}$ being the energy reported by an instrument and E the actual energy of each single event. The apparent spectral model we aim to fit for a single instrument would then be:

$$\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}\tilde{E}} = \frac{\mathrm{d}\phi}{\mathrm{d}E}\frac{\mathrm{d}E}{\mathrm{d}\tilde{E}} = \phi_0 \left(\frac{E/(1+z)}{E_0}\right)^{-\Gamma+\beta\log_{10}\left(\frac{E/(1+z)}{E_0}\right)} \left(\frac{1}{1+z}\right) \tag{4}$$

and the overall joint likelihood is modified in:

$$\mathcal{L}(\mathbf{\Lambda}|\mathbf{D}) = \prod_{i=1}^{n_{\mathrm{instr}}} \mathcal{L}_i(\mathbf{\Lambda}|\{N_{\mathrm{on},ijk}, N_{\mathrm{off},ijk}\}_{j=1,\ldots,n_{\mathrm{runs}};k=1,\ldots,n_{E'}}) \times \mathcal{N}(z_i; 0, \delta_i^2) \tag{5}$$

where now the energy biases are included in the spectral parameters to be fitted: $\mathbf{\Lambda} = (\phi_0, \Gamma, \beta, z_1, \ldots, z_{n_{\mathrm{instr}}})$ and the energy spectrum $\mathrm{d}\tilde{\phi}/\mathrm{d}\tilde{E}$ in Eq. 4 is used to predict the g_{ijk} via forward folding. As in Eq. 1, i runs over the instruments, j on

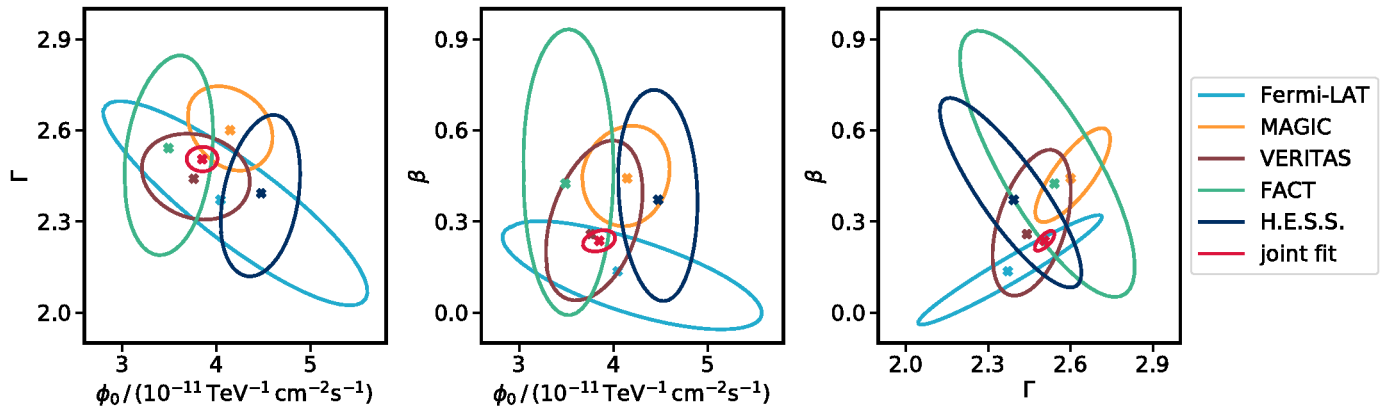

Fig. 3. Likelihood contours corresponding to 68 % probability content for the fitted spectral parameters (ϕ_0, Γ, β), for the likelihood in Eq. 1. Results from the individual instruments and from the joint-fit are displayed.

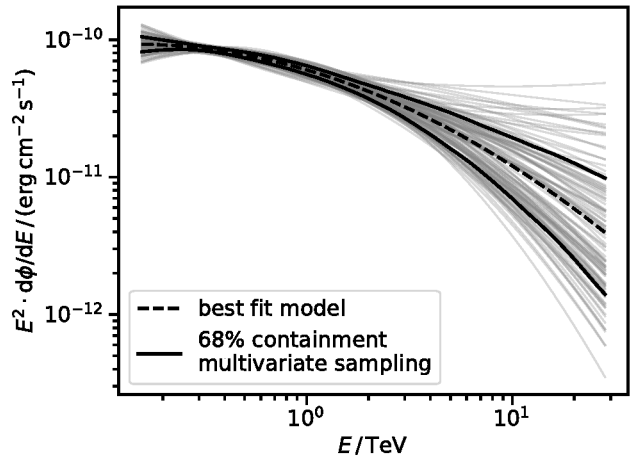

Fig. 4. Error estimation methods for the measured SED using the VERITAS dataset, as example case. The solid black lines display the upper and lower limits of the error band estimated with the multivariate sampling. They represent the 68% containment of 500 spectral realizations (100 displayed as gray lines) whose parameters are sampled from a multivariate distribution defined by the fit results.

the runs and k on the energy bin. The inclusion of the energy biases allows, in addition to the variation of the global spectral parameters ϕ_0, Γ and β (the same for all datasets), also an instrument-dependent energy adjustment (a shift) of the assumed model, through the individual z_i. This shift is not arbitrary: it is, in fact, constrained by its Gaussian distribution with a standard deviation given by the systematic uncertainty on the energy scale provided by the single experiment, δ_i. Hereafter we will refer to this likelihood fit as *stat.+syst. likelihood* that is the generalized version of Eq. 1 (obtainable from Eq. 5 simply fixing all $z_i = 0$). The result of the *stat.+syst. likelihood* joint fit is shown in Fig. 5 in blue against the result of the *stat. likelihood* (Eq. 1) fit in red. We note that in this work we only account for the energy scale systematic uncertainty, as an example of a modified likelihood. A full treatment of the systematic uncertainty goes beyond the scope of this paper. It is possible to reproduce interactively the systematic fit in the online material (`3_systematics.ipynb`).

4. Reproducibility

This work presents a first reproducible multi-instrument gamma-ray analysis, achieved by using the common DL3 data format and the open-source `gammapy` software package. We provide public access to the DL3 observational data, scripts used and obtained results with the GitHub repository mentioned in the introduction, along with a Docker container[9] on DockerHub, and a Zenodo record (Nigro et al. 2018), that provides a Digital Object Identifier (DOI). The user access to the repository hosting data and analysis scripts represents a necessary, but not sufficient condition to accomplish the exact reproducibility of the results. We deliver a `conda`[10] configuration file to build a virtual computing environment, defined with a special care in order to address the internal dependencies among the versions of the software used. Furthermore, since the availability of all the external software dependencies is not assured in the future, we also provide a `joint-crab docker` container, to guarantee a mid-term preservation of the reproducibility. The main results published in this work may be reproduced executing the `make.py` command. This script works as a documented command line interface tool, wrapping a set of actions in different option commands that either extract or run the likelihood minimization or reproduce the figures presented in the paper.

The documentation is provided in the form of Jupyter notebooks. These notebooks can also be run through Binder[11] public service to access via web browser the whole `joint-crab` working execution environment in the Binder cloud infrastructure. The Zenodo joint-crab record, the joint-crab docker container, and the joint-crab working environment in Binder may be all synchronized if needed, with the content present in the joint-crab GitHub repository. Therefore, if eventual improved versions of the joint-crab bundle are needed (i.e. comments from referees, improved algorithms or analysis methods, etc.), they may be published in the GitHub repository and then propagated from GitHub to the other joint-crab repositories in Zenodo, DockerHub and Binder. All these versions would be kept synchronized in their respective repositories.

[9] `https://hub.docker.com/r/gammapy/joint-crab`
[10] `https://conda.io`
[11] `https://mybinder.org/v2/gh/open-gamma-ray-astro/joint-crab/master?urlpath=lab/tree/joint-crab`

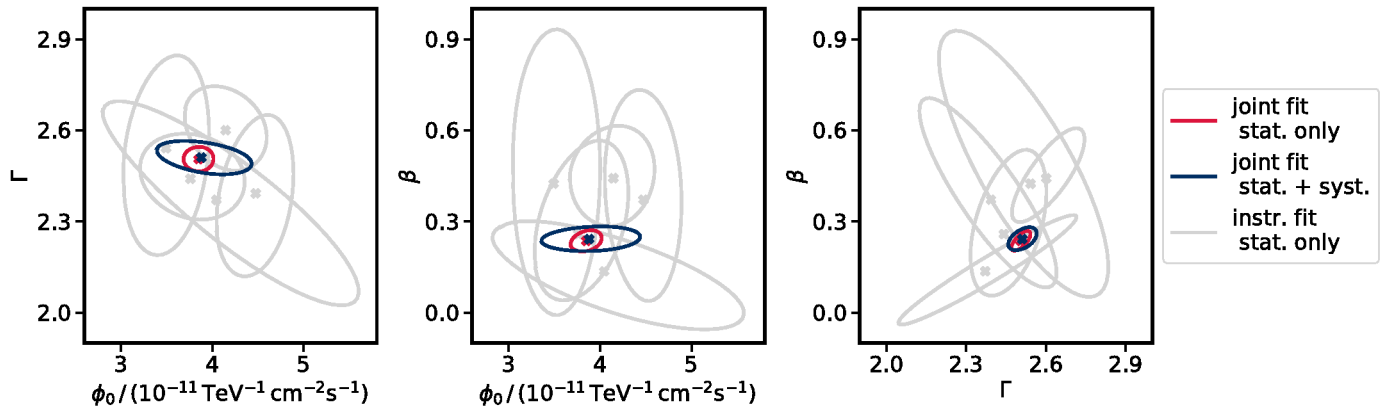

Fig. 5. Likelihood contours corresponding to 68 % probability content for the fitted spectral parameters (ϕ_0, Γ, β), for the likelihood in Eq. 1 (red) and the likelihood in Eq. 5 (blue). Results from the individual instruments with the likelihood in Eq. 1 are displayed in gray.

5. Extensibility

Another significant advantage of the common-format, open-source and reproducible approach we propose to the VHE gamma-ray community is the possibility to access the ON and OFF events distributions and the IRFs, i.e. the results of the spectrum extraction, saved in the OGIP spectral data format[12] (they can be interactively accessed in `1_data.ipynb`). This would allow to perform a maximum likelihood fit to any assumed spectral model $\mathrm{d}\phi/\mathrm{d}E$, that is otherwise impossible. This is of crucial interest for researchers not associated to experimental collaborations that, having usually access only to the final spectral points (often published with no covariance matrix attached), cannot properly test their theoretical models against the data. In the online material of this work, besides the analytically log-parabola function, we considered also a theoretical Synchrotron-Self Compton radiative model (`4_naima.ipynb`) obtained with the open-source `naima` Python package (Zabalza 2015). This is meant to emphasize, on one hand, the potential of the proposed approach, and, on the other hand, the easy interchange between open-source astronomical Python packages with different functionality.

6. Conclusions

This paper presents a multi-instrument reproducible gamma-ray analysis realized with open-source software. It also contains the first public joint release of data from IACTs. Such data dissemination offers the astronomical community the opportunity to gather knowledge of the VHE analysis techniques while waiting for the forthcoming CTA operations. Furthermore they can also be used in data challenges or coding sprints to improve the status of the current science tools.

On a technical note, the DL3 data producible at the moment allow only for joint analyses of a target source at a given position of the FoV (the Crab nebula in this case): no other potential source in the FoV (or multiple sources) can be analysed given the point-like IRFs computed accordingly with the source position that all the instruments, but H.E.S.S., made available. It is worth noting that even the more exhaustive full-enclosure IRFs format may require further development as the current radial offset dependency does not account for a possible non-azimuthal symmetry of the instrument acceptance (Prandini et al. 2015).

On a more general note, the objective of this publication is also to remark a novel approach to gamma-ray science, summarized through the three essential concepts of: common data-format, open-source software and reproducible results. We illustrate that a common data-format allows naturally multi-instrument analysis. Generating data samples compliant with the prototypical DL3 format defined in the "Data formats for gamma-ray astronomy" forum, we perform, for the first time, a spectral analysis of the Crab nebula using data from *Fermi*-LAT and four currently operating IACTs by using the `gammapy` software package. Open-source software will be a key asset for the upcoming CTA, which, as an open-observatory, will share its observation time and data with the wider astronomical community. Reproducible results are seamlessly achieved once the data and software are publicly available. There are several tools and platforms on the market that can be used for this purpose. In particular, with the on-line material attached to this issue, we show a practical example of how a future gamma-ray publication can be released with long-term solutions. A Git repository suffices for the first period after the publication, whereas a Docker accounts for the eventual loss of maintenance of the software packages needed for the analysis. We also provided some considerations on analysis procedures related to spectral analysis commonly performed in the VHE IACT-related astronomy. We proposed a method for computing error bands on the measured SED based on the sampling of a multivariate distribution, along with a method to account for the systematic uncertainties on the energy scales of different gamma-ray instruments while performing a joint fit of their data. We also pointed-out the advantages of publishing the outputs of the spectrum extraction, i.e. the distribution of the signal and background events and the IRFs (alike the OGIP spectral data in the `joint-crab` repository), instead of the spectral points. Mainly this grants the possibility to successively construct a likelihood using an arbitrary theoretical spectral model.

Acknowledgements. This work was supported by the Young Investigators Program of the Helmholtz Association, by the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project C3 and by the European Commision through the ASTERICS Horizon2020 project (id 653477). We would like to thank the H.E.S.S., MAGIC, VERITAS and FACT collaborations for releasing the data that were used. This work made use of `astropy`

[12] `https://gamma-astro-data-formats.readthedocs.io/en/latest/spectra/ogip/index.html`

(Astropy Collaboration 2013) and sherpa (Freeman et al. 2001). The authors are indebted to Abelardo Moralejo and Hans Dembinski for their useful suggestions on the statistical and systematic uncertainty estimation. We are grateful to the anonymous referee for improving the paper with his helpful comments.

References

Abdalla, A. et al. 2018, ArXiv e-prints [arXiv:1810.04516]
Acharya, B. S. et al. 2013, Astroparticle Physics, 43, 3
Aharonian, F. et al. 2001, A&A, 370, 112
Aharonian, F. et al. 2004, ApJ, 614, 897
Aharonian, F. et al. 2006, A&A, 457, 899
Albert, J. et al. 2008, ApJ, 674, 1037
Aleksić, J. et al. 2012, Astroparticle Physics, 35, 435
Aleksić, J. et al. 2015, Journal of High Energy Astrophysics, 5, 30
Aleksić, J. et al. 2016, Astroparticle Physics, 72, 61
Aleksić, J. et al. 2016, Astroparticle Physics, 72, 76
Anderhub, H. et al. 2013, Journal of Instrumentation, 8, P06008
Astropy Collaboration. 2013, A&A, 558, A33
Atwood, W. B. et al. 2009, ApJ, 697, 1071
Ballet, J. et al. 2018, Fermi-LAT 8-year Source List, available at https://fermi.gsfc.nasa.gov/ssc/data/access/lat/fl8y/FL8Y_description_v7.pdf
Barlow, R. J. 2017, ArXiv e-prints [arXiv:1701.03701]
Berge, D. et al. 2007, A&A, 466, 1219
Biland, A. et al. 2014, Journal of Instrumentation, 9, P10012
Brun, R. & Rademakers, F. 1997, Nuclear Instruments and Methods in Physics Research A, 389, 81
Cogan, P. et al. 2008, in Proceedings, 30th International Cosmic Ray Conference (ICRC2007), Merida, Mexico, July 2007, Vol. 3, 1385–1388
Conrad, J. et al. 2003, Phys. Rev. D, 67, 012002
Contreras, J. L. et al. 2015, in Proceedings, 34th International Cosmic Ray Conference (ICRC2015): The Hague, The Netherlands, July 30 - August 6, 2015, Vol. 34, 960
Davis, J. E. 2001, ApJ, 548, 1010
de Naurois, M. & Mazin, D. 2015, Comptes Rendus Physique, 16, 610
Deil, C. et al. 2017a, in Proceedings, 35th International Cosmic Ray Conference (ICRC2017): Busan, South Korea, July 2017, Vol. 34, 766
Deil, C. et al. 2017b, AIP Conference Proceedings, 1792, id.070006
Deil, C. et al. 2018, Data formats for gamma-ray astronomy - version 0.2, doi: 10.5281/zenodo.1409831, https://doi.org/10.5281/zenodo.1409831
Dembinski, H. et al. 2017, in Proceedings, 35th International Cosmic Ray Conference (ICRC2017): Busan, South Korea, July 2017, Vol. 35, 533
Dickinson, H. & Conrad, J. 2013, Astroparticle Physics, 41, 17
Donath, A. et al. 2015, in Proceedings, 34th International Cosmic Ray Conference (ICRC2015), Vol. 34, 789
Fomin, V. P. et al. 1994, Astroparticle Physics, 2, 137
Freeman, P. et al. 2001, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4477, Astronomical Data Analysis, ed. J.-L. Starck & F. D. Murtagh, 76–87
Funk, S. 2015, Annual Review of Nuclear and Particle Science, 65, 245
Hinton, J. A. et al. 2004, New A Rev., 48, 331
Holder, J. et al. 2006, Astroparticle Physics, 25, 391
Kieda, D. B. et al. 2013, in Proceedings, 33rd nternational Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013
Meyer, M. et al. 2010, A&A, 523, A2
Nigro, C. et al. 2018, The joint-crab bundle, doi: 10.5281/zenodo.2381863, https://zenodo.org/record/2381863
Nöthe, M. et al. 2018, ArXiv e-prints, arXiv:1806.01542
Perkins, J. S. et al. 2009, in Proceedings, 2nd Fermi Symposium, Washington, USA, 2-5 November 2009
Piron, F. et al. 2001, A&A, 374, 895
Prandini, E. et al. 2015, in 34th International Cosmic Ray Conference (ICRC2015), Vol. 34, 721
Wells, D. C. et al. 1981, Astronomy and Astrophysics Supplement Series, 44, 363
Zabalza, V. 2015, in Proceedings, 34th International Cosmic Ray Conference (ICRC2015): The Hague, 2015, Vol. 34, 922
Zanin, R. et al. 2013, in Proceedings, 33th Internation Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013, Vol. 34, 773

[1] DESY, D-15738 Zeuthen, Germany
[2] Max-Planck-Institut für Kernphysik, P.O. Box 103980, D 69029 Heidelberg, Germany
[3] Landessternwarte, Universität Heidelberg, Königstuhl, D 69117 Heidelberg, Germany
[4] Instituto de Astrofísica de Andalucía - CSIC, Glorieta de la Astronomía s/n, 18008 Granada, Spain
[5] Unidad de Partículas y Cosmología (UPARCOS), Universidad Complutense, E-28040 Madrid, Spain
[6] APC, AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, 10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
[7] TU Dortmund, Astroteilchenphysik E5b, 44227 Dortmund, Deutschland
[8] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA
[9] Physics Department, McGill University, Montreal, QC H3A 2T8, Canada
[10] Institut de Astrofísica d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology (BIST), E-08193 Bellaterra (Barcelona), Spain
[11] LUTH, Observatoire de Paris, PSL Research University, CNRS, Université Paris Diderot, 5 Place Jules Janssen, 92190 Meudon, France
[12] IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France