

FACT - First Energy Spectrum from a SiPM Cherenkov Telescope

**F. Temme^{a*}, M. L. Ahnen^b, M. Balbo^c, M. Bergmann^d, A. Biland^b, C. Bockermann^e,
T. Bretz^b, K. A. Brügge^a, J. Buss^a, D. Dorner^d, S. Einecke^{a*}, J. Freiwald^a,
C. Hempfling^d, D. Hildebrand^b, G. Hughes^b, W. Lustermann^b, K. Mannheim^d,
K. Meier^d, K. Morik^e, S. Müller^b, D. Neise^b, A. Neronov^c, M. Nöthe^a,
A.-K. Overkemping^a, A. Paravac^d, F. Pauss^b, W. Rhode^a, J. Thaele^a, S. Toscano^c,
P. Vogler^b, R. Walter^c, and A. Wilbert^d**

^aTU Dortmund, Experimental Physics 5

Otto-Hahn-Str. 4, 44221 Dortmund, Germany

^bETH Zurich, Institute for Particle Physics

Otto-Stern-Weg 5, 8093 Zurich, Switzerland

^cUniversity of Geneva, ISDC Data Center for Astrophysics

Chemin d'Ecogia 16, 1290 Versoix, Switzerland

^dUniversität Würzburg, Institute for Theoretical Physics and Astrophysics

Emil-Fischer-Str. 31, 97074 Würzburg, Germany

^eTU Dortmund, Chair of Artificial Intelligence

Otto-Hahn-Str. 12, 44221 Dortmund, Germany

E-mail: fabian.temme@tu-dortmund.de, sabrina.einecke@tu-dortmund.de

The First G-APD Cherenkov Telescope (FACT) is an Imaging Air Cherenkov Telescope located on the Canary Island of La Palma. It is the first of its kind which uses Geiger-mode Avalanche Photo Diodes (G-APDs) as photosensors to detect the Cherenkov radiation emitted from secondary particles in a high-energy gamma-ray air shower. A new analysis chain has been developed using modern data mining methods and unfolding techniques to obtain the energy spectrum of an observed source. This analysis chain has been applied to data of the Crab Nebula, the so called "standard candle" in Cherenkov astronomy. Here, the individual steps are described and results of this application are reported providing the energy spectrum and light curve.

The 34th International Cosmic Ray Conference,

30 July – 6 August, 2015

The Hague, The Netherlands

*Corresponding authors

1. Introduction

FACT (First G-APD Cherenkov Telescope) [5] is the first Imaging Air Cherenkov Telescope which uses Geiger-mode Avalanche Photo Diodes (G-APDs) as photo sensors. It is located at the Observatorio del Roque de los Muchachos on the Canary Island of La Palma (Spain) and since October 2011 it observes very high energy gamma-ray sources in the Northern hemisphere by detecting air showers produced in the atmosphere by very high energy gamma-rays.

FACT is installed on the mount of the former HEGRA CT3 telescope and is equipped with 30 refurbished hexagonal mirrors of the HEGRA CT1 telescope which add up to a total reflective surface of 9.51 m². The camera consists of 1440 pixels yielding a total field-of-view of 4.5° and each of these pixels consists of a solid light concentrator and a Silicon Photomultiplier (SiPM) hosting 3600 G-APDs (Hamamatsu MPPC S10362-33-50C). The novel SiPM photo sensors are mechanically and electronically robust, are capable of single photon counting, have a low operational voltage, a high photon detection efficiency and cheap production costs. These promising properties make them particularly suitable for applications in the ground-based Cherenkov gamma-ray astronomy. More details concerning the performance of this novel technique are given in [6].

The analysis chain starts with the preprocessing of the raw data. For this task, the data analysis tool FACT-Tools [9] has been developed. It performs the calibration of raw data, an extraction of the registered Cherenkov photon pulses, a cleaning process, and calculates several parameters describing the measured image of the air shower. FACT-Tools is embedded in the *streams* framework [7] which has been developed in cooperation with the department of computer science at the TU Dortmund. The image parameters calculated by FACT-Tools are used to perform an energy estimation of the primary particle and a suppression of background showers induced by charged cosmic rays and showers induced by gamma-rays coming from another direction as the source. In principle, both steps could be conducted within the data mining framework RapidMiner [11] (as well developed at the TU Dortmund), but so far the RapidMiner is only used for the background suppression. The RapidMiner framework supports not only the application of a large variety of multivariate methods for classification tasks, but also the necessary validation of a given model. Subsequently, the resulting data set is unfolded using the software TRUEE [12] to obtain the differential energy spectrum of the observed source. Figure 1 shows an overview of the described analysis chain.

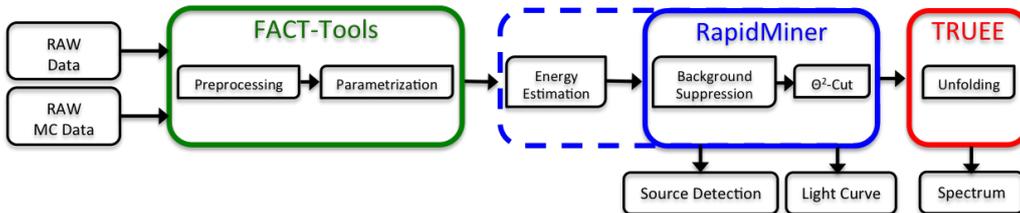


Figure 1: Schematic flow diagram of the FACT analysis chain.

The Crab Nebula, a nearby pulsar wind nebula, is the brightest source in high energy gamma-rays with no detected variability so far. Therefore it serves as the "standard candle" in very high energy gamma-ray astronomy to evaluate and compare the performance of instruments. The different steps of the analysis chain will be presented and the Crab Nebula energy spectrum measured by FACT will be shown for the first time.

2. Data Sample

In order to evaluate the performance of the FACT telescope and its corresponding analysis chain, data of the Crab Nebula taken between October 2013 and February 2014 is used. By a dedicated data selection, the final sample contains only data taken during good conditions, meaning a zenith angle below 30° , good weather, no moon light and a stable trigger rate. This data have been taken in the so-called wobble mode, i.e. the telescope points at the source (ON) with a specific offset, which allows to estimate the background (OFF) under the same conditions at the same time.

Dedicated Monte Carlo simulations (MCs) are necessary for the analysis of these data. They are generated using the software CORSIKA [10] to model the air shower and the software ceres provided within the MARS - CheObs ed. package [8] to model the telescope's response. The MC set includes 12 million air showers induced by gamma-rays in an energy regime between 200 GeV and 50 TeV and about 780 million air showers induced by protons in an energy range between 100 GeV and 200 TeV. Both energy spectra follow a power law with an index of -2.7.

3. Preprocessing and Parameterization

The first steps of the analysis chain are the preprocessing and parameterization of the raw data. Therefore, the data analysis tool FACT-Tools is used, which calibrates the raw data, extracts the number of Cherenkov photons and identifies the image of the measured air shower by applying an image cleaning algorithm. Subsequently, the image is parameterized and the calculated parameters are applicable in the following steps of the analysis chain. More details about FACT-Tools can be found in [9].

Figure 2 shows the distribution of two exemplary parameters calculated by FACT-Tools for the proton MCs and the measured data of the Crab Nebula. Note, that no background suppression has been applied yet to this Crab Nebula data and contains mostly protons due to the much larger fraction (ratio $\sim 1:1000$) of proton showers compared to showers induced by gamma-rays and other background showers. This justifies a comparison between proton MCs and measured Crab Nebula data. The comparison shows an agreement in the general shape of the distribution. However, the rates in the MCs are slightly higher than in measured data.

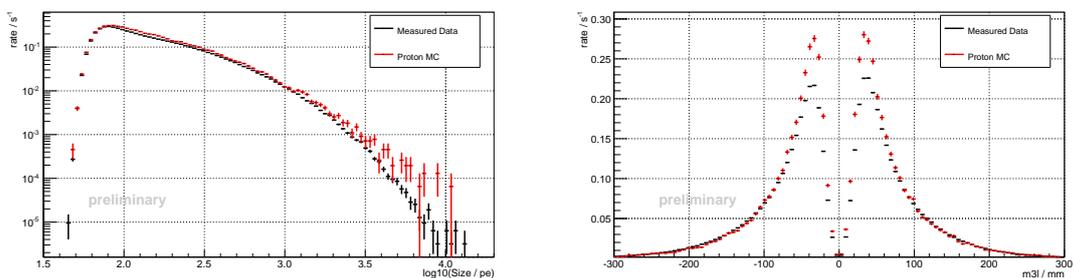


Figure 2: Comparison of parameter distributions normalized to observation time, using the example of the total light content of the shower (*left*) and the third moment along the major axis of the shower ellipse (*right*). While the general shape of the distributions is reproduced by the MCs, the rates in the MCs are slightly higher than in measured data.

4. Energy Estimation

Since it is not possible to measure the energy of the incoming particle directly, it is necessary to estimate it based on the measured image parameters which feature a correlation to the sought-after energy of different extent. For this energy estimation a random forest regression is performed with the open source machine learning library `scikit-learn` [13] resulting in a new parameter called reconstructed energy (E_{rec}). The correlation between the simulated energy (E_{MC}) and this new parameter E_{rec} is displayed in Figure 3. It shows a clear correlation but an unambiguous conclusion from the reconstructed to the true energy is not possible.

The performance of the energy estimator is evaluated by calculating the bias and resolution caused by the telescope and the analysis. Hence, a gaussian function is fitted to the distribution of $(E_{\text{rec}} - E_{\text{MC}})/E_{\text{MC}}$ for individual energy bins. The mean and the standard deviation of this function estimates the bias and the resolution. The results are visualized in Figure 3.

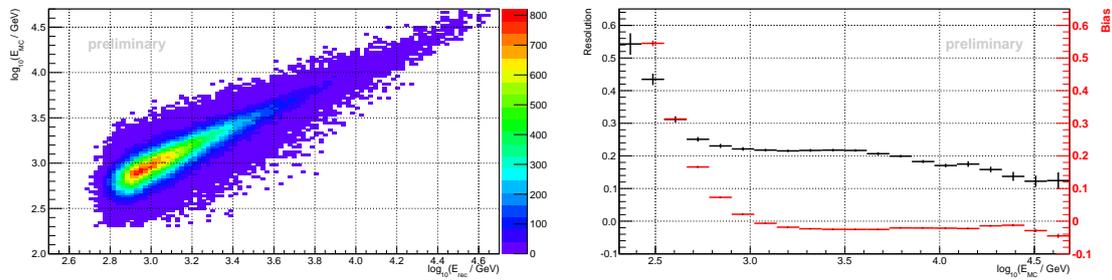


Figure 3: *Left:* Correlation between true energy (E_{MC}) and reconstructed energy (E_{rec}) by evaluating gamma MCs. *Right:* Energy resolution and bias obtained by evaluating gamma MCs.

As the random forest regressor is trained with gamma MCs, the variance of the confidence of the ensemble of decision trees for proton MCs is much larger than for gamma MCs. This can be seen in Figure 4. Therefore, the variance of the energy estimator is well suited during the background suppression.

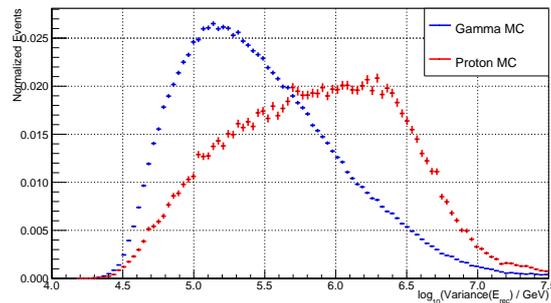


Figure 4: Variance of E_{rec} of each particular tree in the random forest regression for proton and gamma MCs, normalized to the number of events. This parameter is well suited during the background suppression.

5. Background Suppression

In order to suppress the dominant background showers induced by charged cosmic rays or gamma-rays not coming from the direction of the source, a random forest classification is performed within the RapidMiner. Therefore a random forest with 100 trees, 45 total features and 7 features per node is trained with 50000 signal and 50000 background events.

After the classification model has been built, it is applied to the measured data and to disjoint MCs and a confidence is calculated for each event describing the probability of an event to belong to the positive (here: signal) class. In order to obtain data sets for the following analysis steps, a cut in this confidence has to be applied. The choice of values for this cut depends on the particular purpose of the data set to be generated. They are evaluated by calculating quality parameters on an independent MC test sample, validated by a 10 fold bootstrap validation. By way of example, Figure 5 shows the gamma and proton efficiency for different confidence cuts. To evaluate the values for the cuts, the F_β score [14] is used:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{efficiency}}{(\beta^2 \cdot \text{precision}) + \text{efficiency}} \quad (5.1)$$

The F_β score measures the quality of the classification with respect to the importance of the precision compared to the efficiency by attaching a factor β . This factor can be tuned according to the particular intended use, e.g. a source detection has different requirements to the data set as the determination of an energy spectrum. Especially in the case of highly unbalanced data, where the minority class is more important, the F_β score is well-suited. For the source detection $\beta = \frac{1}{20}$ and for the determination of the energy spectrum $\beta = \frac{1}{6}$ are chosen. The confidence cut is chosen according to the maximum of F_β score.

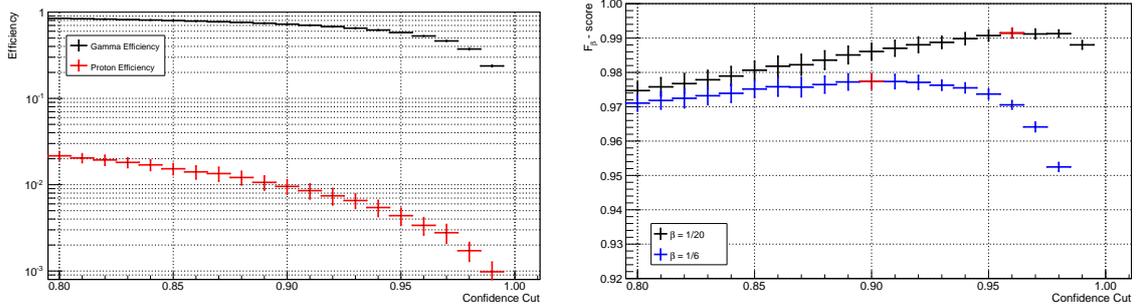


Figure 5: *Left*: Gamma and proton efficiency for different confidence cuts. *Right*: $F_{\frac{1}{20}}$ and $F_{\frac{1}{6}}$ score evaluated with a 10-fold validation on an independent MC set. According to the maximum of the distributions, the confidence cuts for the source detection and for the unfolding are chosen. The chosen F_β scores for the particular confidence cuts are marked in red.

6. Source Detection

As stated in the previous section, the source detection requires specific cuts to obtain the best suitable data set for this purpose. A confidence cut of 0.96 is chosen to perform the detection of the Crab Nebula. The application of the random forest model on the Crab Nebula data yields to a clear signal of the source. Figure 6 shows the distribution of θ^2 for ON and OFF data. The source is detected with a significance of 43.54σ in 87.6 hours.

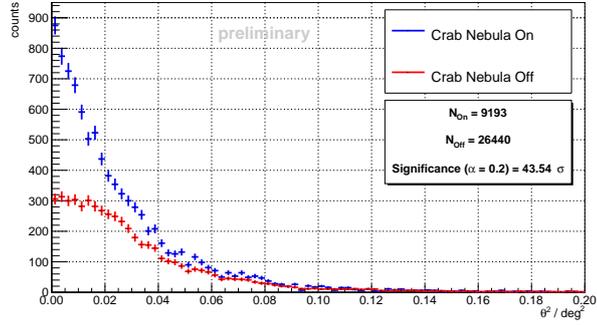


Figure 6: θ^2 -Plot for Crab Nebula data with a confidence cut of 0.96. Note, that θ^2 is used in the classification model as well, resulting in a non-flat OFF distribution.

7. Energy Spectrum

As shown in section 4, an unambiguous conclusion from the reconstructed to the true energy is not possible, whereas an unfolding procedure has to be conducted. The software TRUEE provides a well tested algorithm with the capability to use up to three image parameters during the unfolding fit. This fit is based on a regularized likelihood fit and unphysical fluctuations can be suppressed by the application of a Tikhonov regularization. Furthermore, it provides different modes for the validation of e.g. the chosen settings and several quality checks of the unfolding. A more detailed description of TRUEE, its implementation and capabilities are presented in [12].

The unfolding step requires slightly different cuts on the confidence (0.9) as for the source detection. By choosing the reconstructed energy, the light concentration in the shower image and the zenith angle for the unfolding fit, an energy spectrum of the Crab Nebula has been obtained and is shown in Figure 7. The slightly higher flux could be explained due to the higher rates seen in MCs (compare section 3). However, the energy spectrum of the Crab Nebula obtained with the FACT telescope is consistent with previous measurements of this source performed with other IACTs.

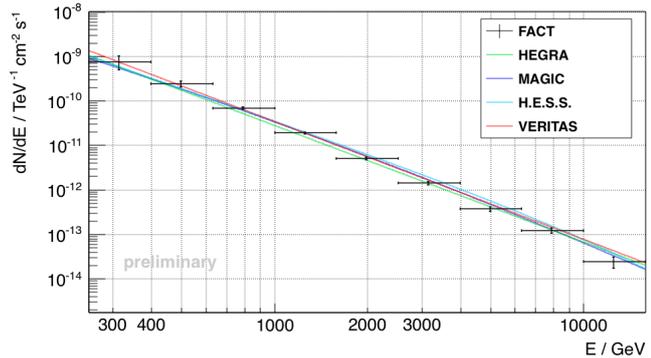


Figure 7: Unfolded energy spectrum of the Crab Nebula obtained with FACT compared to results from other telescopes: HEGRA [2], MAGIC [3], H.E.S.S. [1] and VERITAS [4]. The vertical error bars indicate only statistical uncertainties. The slightly higher flux could be explained due to the higher rates seen in MCs.

One of the quality checks of the unfolding is the comparison of the distribution of an image parameter not used during the unfolding fit between the measured data and the MC simulation reweighted with the unfolding result. The result of this comparison is shown in Figure 8 using the example of the parameter *Length*. The agreement of both distributions indicates a proper performance of the unfolding.

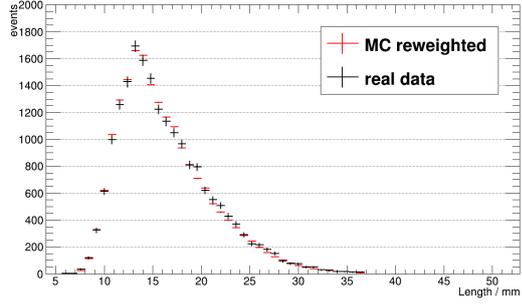


Figure 8: Distributions of the image parameter *Length* for measured data and MC simulations reweighted with the unfolding result. The agreement of the distributions indicates a proper unfolding.

By means of the unfolded energy spectrum, an integrated flux has been calculated as a function of time (Figure 9). The daily flux points are consistent within the uncertainties with an expected constant flux. The average integrated flux is consistent with measurements of other IACTs, taking into account the uncertainties of this measurement as well as the ones of the other experiments. The slightly higher flux could be due to the higher rates seen in MCs (compare section 3).

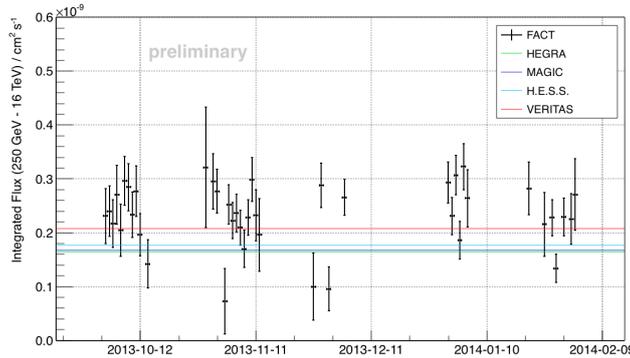


Figure 9: Lightcurve of the Crab Nebula between September 2013 and February 2014 measured with FACT and compared to other experiments as in Figure 7. The slightly higher flux could be due to the higher rates seen in MCs. The daily flux points are consistent with a constant flux.

8. Conclusions

In this work the analysis chain of the FACT telescope including modern data mining methods and unfolding techniques has been described and the results of its application on Crab Nebula data have been presented. The source has been detected with a significance of about 44σ in 87.6 h and the lightcurve as well as the energy spectrum between 250 GeV and 16 TeV are consistent with measurements of other experiments. The validation of all analysis steps has been important to achieve these results.

Acknowledgment The important contributions from ETH Zurich grants ETH-10.08-2 and ETH-27.12-1 as well as the funding by the German BMBF (Verbundforschung Astro- und Astroteilchenphysik) and HAP (Helmholtz Alliance for Astroparticle Physics) are gratefully acknowledged. We are thankful for the very valuable contributions from E. Lorenz, D. Renker and G. Viertel during the early phase of the project. We thank the Instituto de Astrofísica de Canarias allowing us to operate the telescope at the Observatorio del Roque de los Muchachos in La Palma, the Max-Planck-Institut für Physik for providing us with the mount of the former HEGRA CT 3 telescope, and the MAGIC collaboration for their support. Part of this work is supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Research-Constrained Analysis", project C3.

References

- [1] F. Aharonian et al. *Observations of the Crab nebula with HESS*. In: *A&A* 457.3 (2006).
- [2] F. Aharonian et al. *The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA Stereoscopic Air Cerenkov Telescopes*. In: *ApJ* 614.897 (2004).
- [3] J. Aleksic et al. *The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula*. In: *ApJ* in press (2015).
- [4] E. Aliu et al. *A search for enhanced very high energy gamma-ray emission from the 2013 March Crab Nebula flare*. In: *ApJ* 781.L11 (2014).
- [5] H. Anderhub et al. *Design and operation of FACT – the first G-APD Cherenkov telescope*. In: *JINST* 8.06 (2013), P06008.
- [6] A. Biland et al. *Calibration and performance of the photon sensor response of FACT — the first G-APD Cherenkov telescope*. In: *JINST* 9.10 (2014), P10012.
- [7] C. Bockermann and H. Blom. *The Streams Framework*. Tech. rep. 5. TU Dortmund, Dec. 2012.
- [8] T. Bretz and D. Dorner. *MARS - CheObs goes Monte Carlo*. In: *Proceedings of the 31st ICRC*. 2009.
- [9] K. Brügge et al. *FACT-Tools: Streamed Real-Time Data Analysis*. In: *these proceedings*. 1135. 2015.
- [10] D. Heck et al. *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers,* Report FZKA 6019, Forschungszentrum Karlsruhe. Tech. rep. 1998.
- [11] I. Mierswa et al. *YALE: Rapid Prototyping for Complex Data Mining Tasks*. In: *Proceedings of the 12th ACM SIGKDD*. 2006.
- [12] N. Milke et al. *Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics*. In: *NIM A* 697 (2013), pp. 133–147.
- [13] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [14] M. Sokolova et al. *Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation*. In: *Proceedings of the 19th Australian Joint Conference on AI: Advances in Artificial Intelligence*. 2006.