Asterics
Astronomy ESFRI & Research Infrastructure Cluster

ASTERICS - H2020 - 653477

Technology Benchmark Report D-GEX Work Package

ASTERICS GA DELIVERABLE: D3.18

Document identifier:	ASTERICS-D3.18.docx
Date:	11 May 2019
Work package:	WP3 OBELICS
Lead partner:	ASTRON
Document status:	Final
Dissemination level:	Public
Document link:	https://www.asterics2020.eu /documents/ASTERICS-D3.18.pdf

Abstract

In this report, we present results and work that have been carried out under task 3.2 of OBELICS WP, D-GEX, with focus on benchmarking, but also including the rationale for the software development and a description of the functionality of the packages. The contributions presented are all new and serve as an overview of the work done in the work package in the last two years 2017-2019.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/WP	Date
From	José Luis Contreras	UCM	23-05-2019
Author(s)	Thomas Vuillaume (LAPP) Peter Hague (UCAM) Paschal Coyle (CCPM) Lea Jouvin (IFAE) J. Rosado (UCM) S. Van der Tool (ASTRON) J.L. Contreras (UCM)	LAPP UCAM CCPM IFAE UCM ASTRON UCM	
Reviewed by	Tammo Jan Dijkema	ASTRON	
Approved by	AMST		24-05-2019

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	10-03-2019	First scheme	J.L. Contreras, UCM
2	30-04-2019	Contribution from several institutions	
3	20-05-2019	First complete Draft	J.L. Contreras
4	24-05-2019	Revise version	R. Van der Meer

IV. APPLICATON AREA

This document is a formal deliverable for the GA of the project, applicable to all members of the ASTERICS project, beneficiaries and third parties, as well as its collaborating projects.

V. TERMINOLOGY

ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster
CERN	European Organization for Nuclear Research
CTA	Cherenkov Telescope Array
E-ELT	European Extremely Large Telescope
ESFRI	European Strategy Forum on Research Infrastructures
HDF5	Hierarchical Data Format version 5
H.E.S.S.	High Energy Stereoscopy System

HiSCORE	Hundred *I Square-km Cosmic ORigin Explorer
KM3NeT	Cubic Kilometre Neutrino Telescope
LOFAR	The Low Frequency Array
LSST	The Large Synoptic Survey Telescope
MAGIC	Major Atmospheric
OBELICS	Observatory E-environments LInked by common ChallengeS
SCADA	Supervisory Control And Data Acquisition
SKA	The Square Kilometre Array
VLBI	Very Long Baseline Interferometry

A complete project glossary is provided at the following page:
http://www.asterics2020.eu/glossary/

VI. PROJECT SUMMARY

ASTERICS (Astronomy ESFRI & Research Infrastructure Cluster) aims to address the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities (SKA, CTA, KM3Net & E-ELT). It brings together for the first time, the astronomy, astrophysics and particle astrophysics communities, in addition to other related research infrastructures. The major objectives of ASTERICS are to support and accelerate the implementation of the ESFRI telescopes, to enhance their performance beyond the current state-of-the-art, and to see them interoperate as an integrated, multi-wavelength and multi-messenger facility. An important focal point is the management, processing and scientific exploitation of the huge datasets the ESFRI facilities will generate. ASTERICS will seek solutions to these problems outside of the traditional channels by directly engaging and collaborating with industry and

specialised SMEs. The various ESFRI pathfinders and precursors will present the perfect proving ground for new methodologies and prototype systems. In addition, ASTERICS will enable astronomers from across the member states to have broad access to the reduced data products of the ESFRI telescopes via a seamless interface to the Virtual Observatory framework. This will massively increase the scientific impact of the telescopes, and greatly encourage use (and re-use) of the data in new and novel ways, typically not foreseen in the original proposals. By demonstrating cross-facility synchronicity, and by harmonising various policy aspects, ASTERICS will realise a distributed and interoperable approach that ushers in a new multi-messenger era for astronomy. Through an active dissemination programme, including direct engagement with all relevant stakeholders, and via the development of citizen scientist mass participation experiments, ASTERICS has the ambition to be a flagship for the scientific, industrial and societal impact ESFRI projects can deliver.

Table of Contents

1. Introduction

This document is focuses on the work carried out in the D-GEX task in the OBELICS work package of the ASTERICS project in the last two years of the project 2017-2019. It presents the main lines of work developed within D-GEX by the different participating institutions. Nearly all the institutions contributing to the D-GEX subpackage present work: Laboratorie d´Annecy de Physique des Particules (LAPP), University of Cambridge (UCAM), Centre de Physique des Particules de Marseille, Institut de Fisica d´Altes Energies (IFAE), Universidad Complutense de Madrid (UCM) and the Netherlands Institute for Radioastronomy (ASTRON). Those not present here have either described the work in other deliverables or/and published with references to ASTERICS.

The deliverable was originally conceived as a set of benchmarks on the technologies used in the work package, but the institutes involved have decided to be more ambitious and report also on their full work, while also mentioning the benchmarks performed.

One of the things discovered during these years of work has been the difficulty to build walls between the different aspects of the Data acquisition and analysis chain. The three tasks defined inside OBELICS, representing three different levels in the data process road, have often influenced each other and sometimes it has been difficult to assign work to one or the other. Therefore, this deliverable should be considered as part of a set of three, which, in addition, contains deliverables 3.19 from D-INT, and 3.20 from D-ANA.

1. Polynomial data compression (LAPP/CNRS)

The new generation research experiments will introduce a huge data surge to a continuously increasing data production by current experiments. This data surge necessitates efficient compression techniques. These compression techniques must guarantee an optimum trade-off between compression rate and the corresponding compression /decompression speed ratio without affecting the data integrity. This work presents a lossless compression algorithm to compress physics data generated by Astronomy, Astrophysics and Particle Physics experiments.

The algorithm relies on the fact that in many cases of digitized data, the range of covered values by the data is lower than 232 corresponding to the maximal value to be stored as an unsigned integer. In this case, several pieces of data can be stored in a single unsigned int. The calculation of the stored values is done using a polynomial approach where the power of the base is given by the values range. The figure 1 summarizes the algorithm. The upper line represents the data (different colours for different values). In the second line, the orange blocks represent the changes between the different values to compress. The last line shows the compressed data (as they are stored). First, the minimum value of the data, next, the base b = max − min + 1, which defines the data variations set, Z/bZ and finally the data variations. Several data can be stored in the same unsigned int and only the changes between the data are stored. The common parameters like the range of the data (minimum and maximum or compression base) are stored only once.

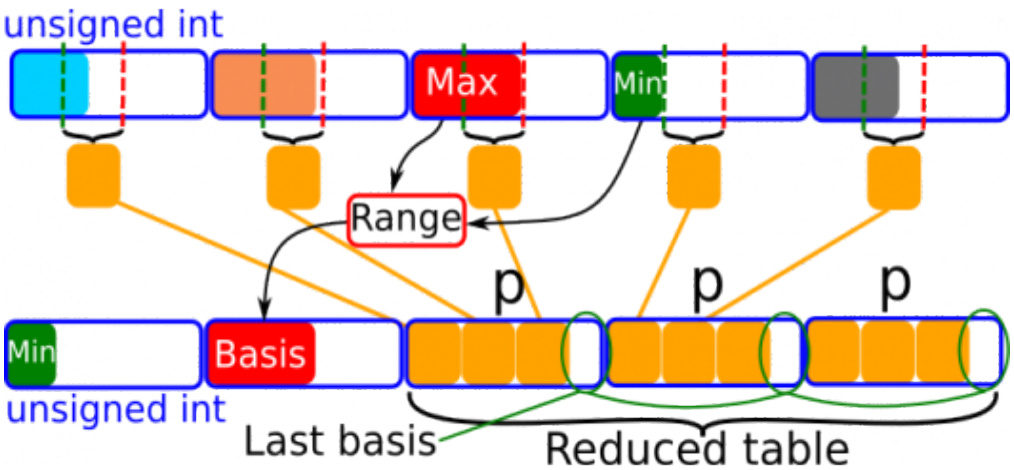

Figure 1: Illustration of the polynomial compression. Credit: P. Aubert, LAPP, CNRS

The compression method has then been further improved by storing slices of data in the unused space at the end of each packed unsigned int.

The developed algorithms have been tuned and tested on a real use case: the next generation ground-based high-energy gamma ray observatory, Cherenkov Telescope Array (CTA), requiring important compression performance. Stand-alone, the proposed compression method is very fast and reasonably efficient. Alternatively, applied as pre-compression algorithm, it can accelerate common methods like LZMA, keeping close performance in terms of compression ratio.

This work has been published in [1].

2. BaSC - Bayesian Source Characterisation (UCAM)

BaSC is an advanced Bayesian source finding library that uses a likelihood model that is mathematically proven in terms of visibilities but which can operate using only a dirty map; thus gaining accuracy without large computational expense. Source finding that uses CLEANed maps is subject to the loss of information necessarily caused by that algorithm. In fact, no matter the quality of such algorithms, the best that they can hope to recover is the CLEAN model – not the actual sources. BaSC works on dirty maps, which still have the messier PSF that is produced by interferometers but do not suffer a loss of information. We apply a Markov Chain Monte Carlo (MCMC) method alongside a likelihood function on the map that is proven equivalent to a (much more expensive) function for the visibilities. The result is more accurate source positions and fluxes, better ability to discriminate sources near the level of instrument resolution, and a more transparent provenance for the source list produced.

BaSC is provided as a Python 3 library. Download from the Git repository above and, for the case of Linux type ‘make’ or for Mac type ‘make mac’. Windows is not natively supported at this time. A user manual is included in the repository. This code will continue development at the University of Cambridge once ASTERICS is concluded. Focus will be on including more functionality within BaSC – presently generation of the dirty maps has to be done externally with programs such as CASA, but there is an advantage of being able to process the visibilities directly within BaSC before applying the source finder. Optimisation of the MCMC process is also being looked at; at present, the source finder can take a long time to run in the case of many (>30) sources due to degeneracies in the parameter space. We aim to improve this. Wide field images can cause issues in BaSC due to the change in the PSF across the field. There are a number of ways to address this, which are being investigated

The latest developments on this package time period were concerned with the development and testing of BaSC. The software is now in a useable form, publicly available on github. We

have produced an article [2] that outlines the results of the tests we have done, and confirms that the software is superior at source discrimination tasks to the usual pathway for radio astronomy of CLEAN/SExtractor and approaches the mathematically optimal resolution limit.

We were able to calculate the optimal performance for a constrained task (discriminating between two nearby points) and then construct appropriate testing sets. The design of the experiment was critical to confirming that BaSC did indeed work as expected. Efforts required & impact of these solutions (writing codes, testing codes, bench marking etc).

The results of this work are available in [2].

3. OMGsim and ParamNu (CCPM)

The packages OMGsim and ParamNu have been developed at CCPM for the simulation and analysis of neutrino detector data.

OMGsim: Optical Module Geant4 simulation

OMGsim, Optical module Geant4 simulation, software objective is to provide an easy to use simulation of the optical modules containing PMTs, in particular, multi-PMT optical modules of KM3NeT. An important attention has been given to the simulation of the photocathode, using a dedicated thin layer and complex refraction index to determine photon absorption, reflection and transmission. Photoelectron emission is simulated using Spicer's three-step model.

This simulation will allow you to:

- Easily reproduce precise geometry of the detector inside the simulation (human-readable config files).
- Easily determine a precise geometry of the phototubes and optical modules (human-readable config files).
- Easily determine the different materials and properties of the detector and phototubes (Human-readable config files with CSV tables).
- Test any experiment you would like to reproduce:
    - Laboratory angular acceptance scans,
    - Optical background studies from the optical module glass,
    - ^{40}K decay in water.

More information about the package can be found in reference [3].

ParamNU

ParamNu is a lightweight, ROOT-based neutrino telescope simulation package, initially intended for KM3NeT/ORCA, but generalisable to IceCube and similar experiments. The package enables studies of the experiment's sensitivity to fundamental neutrino properties, including the neutrino mass ordering, mixing angles, squared mass splittings, etc. The package allows to calculate event number distributions and asymmetries, based on parameterized detector response functions.

The detector response is based upon simple, continuous and smooth mathematical functions, allowing this package to be easily shared outside the KM3NeT Collaboration in a way that is independent from the KM3NeT simulation chain. Neutrino oscillation probabilities are calculated using the OscProb package. The sensitivity calculations use the Asimov approach. Systematic uncertainties are incorporated in the calculations as priors, allowing to configure both the neutrino oscillation parameter uncertainties as well as detector-related uncertainties.

A model of the atmospheric neutrino flux and a tabulated set of neutrino-nuclei cross section data are included with the package.

The package is most easily run on a Linux system. The main requirements are: a C++ compiler, ROOT (with Minuit enabled at compilation time), OscProb. Additional requirements include the standard C++ libraries, as well as the libconfig++ and libconfig++-dev packages.

More information about the package can be found in reference [4].

4. Conversion of MAGIC data to the DL3 format (IFAE)

CTA will be the first ground-based gamma-ray telescope array operated as an open observatory with public observer access. Contrary to the current IACT, like H.E.S.S., MAGIC or VERITAS, for which the data and software are mostly private to the collaborations operating the telescopes, the data and software will be public at a certain level of the analysis. This implies strong requirements to define the format and to design the software tools that will allow to perform high level analysis of IACT data. This open format, named DL3, is based on the FITS format and will contain a list of reconstructed gamma-ray events for each observation as well as the associated instrument response function (described in the D3.19 report).

In order to test this format and the new tools, it is necessary to convert the current IACT data into this DL3 format. Those works are major steps towards a common legacy of the

data and vehicle the strength of open and reproducible multi-instrument analysis in gamma-ray astronomy. In MAGIC, this conversion started some years ago with the creation of a MAGIC convertor inside the MAGIC software [5]. Thanks to this converter, we participate to the first joint likelihood analysis of the Crab Nebula using data from different gamma-ray instruments (reported in deliverable D3.19).

We also started to work on a general pipeline to deliver a large dataset of the data, like it was provided by the H.E.S.S. collaboration in September 2018. This work is still ongoing. Moreover, you can have two types of sources in the observations: pointlike or extended. In the case of a pointlike observation (majority of the sources in MAGIC), we use the instrument response function (IRF) named pointlike where you apply a cut on the event direction from the source position to increase the signal to noise ratio when you produce your spectrum. What we started to develop in the general pipeline, is to produce also for each observation what we call full enclosure IRF (with no cut applied on the event direction) and offset dependent field of view instrument response function. Both are needed for morphological and 3D analysis that are a crucial point for the high-level analysis software developed for CTA. We validated this work on some observation of the Crab Nebula in MAGIC taken at different offset from the camera centre. Before using the 3D analysis, we have to develop a background model for each observation in the DL3 converter. This work just started and will be ongoing.

5. Implementation of air-fluorescence emission in CORSIKA code (UCM)

Monte Carlo simulations of cosmic-ray air showers are required to characterize the performance of Cherenkov telescopes, e.g., instrument response functions. CORSIKA [6] is the most used Monte Carlo code in very high-energy astrophysics. It includes the detailed simulation of the Cherenkov light emission [7,8], while the air fluorescence emission, which is within the same wavelength range but less efficient, has not been implemented yet in the official CORSIKA software. Indeed, the fluorescence component has so far been neglected in Cherenkov telescopes. However, next-generation imaging atmospheric Cherenkov telescopes (IACTs) of the CTA Observatory [9] will achieve such a high sensitivity that the small contribution of fluorescence light may be significant in certain circumstances.

The UCM group has developed a code for the fluorescence emission within the CORSIKA framework. The code is thoroughly described in [10], but it essentially generates fluorescence photons along the track of every simulated charged particle as well as at any point where a particle falls below the cut-off. This allows for a simulation of the fluorescence emission with a level of detail as high as that of the Cherenkov emission. Information on fluorescence photons is written in the same way as Cherenkov photons, either in a standard CORSIKA output file or in the dedicated output format for IACTs. As an example, the

obtained spatial distribution of fluorescence photons on the ground for 10 TeV gamma-ray vertical showers is compared to that of Cherenkov photons in Fig. 2.

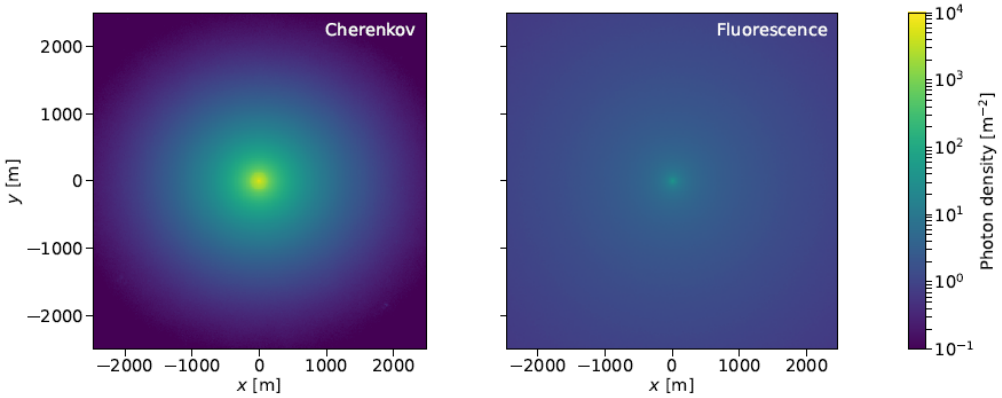

Fig. 2. Spatial distributions of Cherenkov (left) and fluorescence (right) light on ground obtained by the simulation for 10 TeV gamma-ray vertical showers.

The code is planned to be incorporated into upcoming official CORSIKA v7 releases and, later on, in CORSIKA 8 [11]. Meanwhile, a modified CORSIKA code including fluorescence emission is available under request to the UCM group. Besides, the simulation tool simtel array for IACT systems [12] has already been adapted by the developer, allowing for obtaining camera images from Cherenkov and fluorescence light separately, as illustrated in Fig. 3.

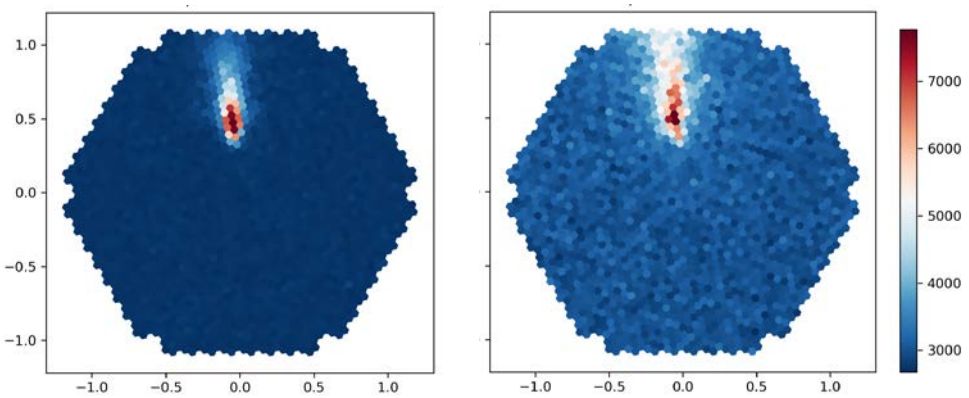

Fig. 3. Simulated IACT images from Cherenkov (left) and fluorescence (right) light of an air shower using the simulation tool *simtel_array*. The fluorescence light has been amplified by a factor of 1000.

This implementation of the fluorescence emission in CORSIKA has been tested against numerical calculations based on a simplified one-dimensional shower development [5,8]. Results show an excellent agreement when considering effects due to the lateral profile of the air showers, as shown in Fig. 4.

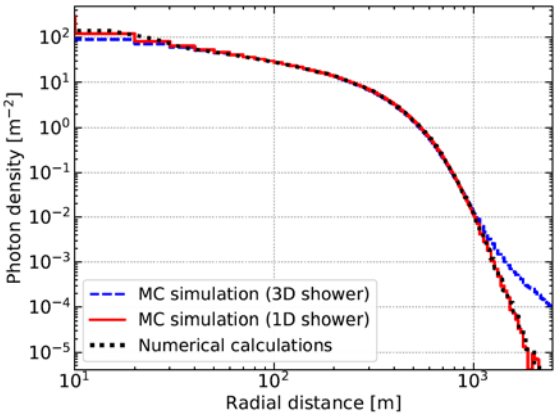

Fig. 4. Validation of the Monte Carlo simulation against numerical calculations. Differences are justified by the effect of the lateral profile of the air shower in the photon density distribution on ground.

The simulation of Cherenkov emission consumes a large fraction of the CPU time spent in CORSIKA. When including the fluorescence emission as well, this CPU time increases by a factor of 5-6 (e.g., from 236 s to 1364 s for 100 gamma showers of 1 TeV, for a 3300 MHz CPU). In addition, the output file size also increases by around 6% with respect to the simulation without producing fluorescence. Although both represent substantial runtime and size increments, it is still feasible to perform dedicated MC productions thanks to the huge computational power available nowadays. Nonetheless, we plan to work on the optimization of the code in future versions.

The code has already been used to evaluate in a systematic way the relevance of the fluorescence radiation in Cherenkov telescopes [11]. This light contribution becomes important when increasing the distance from the shower core impact point, because the Cherenkov light density decreases dramatically while that of the fluorescence light keeps rather constant due to its isotropic emission, as shown in Fig. 2. For IACTs, the fluorescence contamination reaches 5% at about 1000 m. The effect is more prominent for arrays of wide-angle Cherenkov detectors, like HiSCORE [14], where the fluorescence contribution is found to be as high as 45% at a core distance of 1000 m.

6. ARTAMIS: monitoring system for Apertif (ASTRON)

With the APERTIF phased array feed upgrade to the Westerbork Synthesis Radio Telescope, each dish has gained the capability of observing in 40 directions simultaneously. This has also increased the complexity of the system, and thus the complexity of monitoring the system health. We have developed ARTAMIS, a monitoring system for monitoring all of Westerbork, including dish control, weather sensors, observations, dataflow, and the state of external parts of the system like Radio Astron, Galileo, (e)VLBI, etc.

This gives us the capability to see everything we need in one monitoring system. Features include automatic alarms, loading data points into graphs, and looking back in time to spot anomalies. These anomalies can include failing fans, firmware problems, failing dishes, broken PAF elements. Moreover, the software enables us to read out spectrum analyzers, or see an overview of running observations. This gives us the capability to know within seconds if the hardware is capable of delivering what we need, all the correct firmware and software has been loaded for daily operations.

ARTAMIS is built in the WinCC Open Architecture, a SCADA system for visualizing and operating processes, production flows, machines and plants in all lines of business. WinCC is also used in monitoring the CERN facilities. Our ARTAMIS plugin was built using periodic feedback sessions with input from the people building the hardware, enthusiastic astronomers, software engineers, field engineers and other future users to see what needs to be monitored, what can be monitored and who needs what data points to be displayed in what way.

Technically, the input for the monitoring system is taken from multiple data sources like the generic APERTIF message bus, JSON data sources, Cacti, and more.

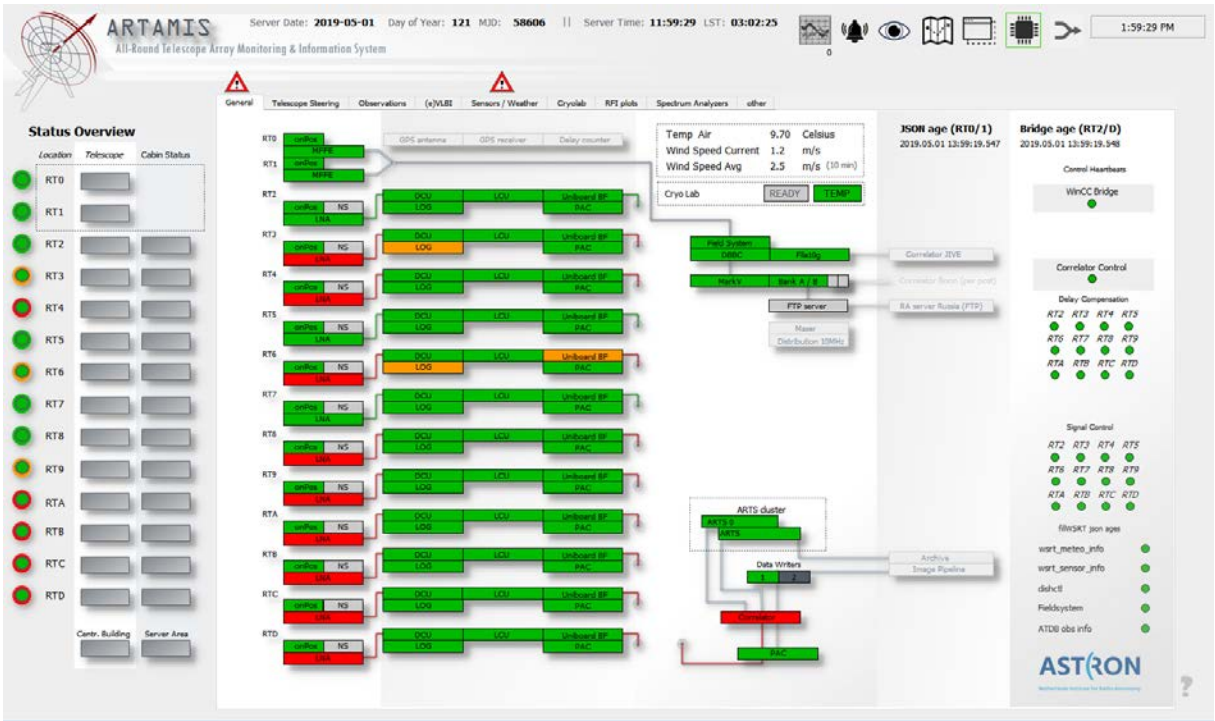

Fig. 5. A screenshot of the ARTEMIS system.

7. MeasurementSet performance testing (ASTRON)

Radio interferometric data is commonly stored in a domain specific data format called 'MeasurementSet' (MS). For the MSv3 project, the current MeasurementSet IO performance has been benchmarked. The MS performance has been compared to HDF5 by writing and reading the same data sizes and types in an HDF5 file. Also, it has been compared to writing raw IO with the dd tool.

Various MS sizes, access patterns, computer systems and other parameters have been used to collect sufficient performance data.

To do the benchmarking two programs have been developed:

- writems to create and fill a number of MeasurementSets (MSs). Various parameters (such as number of fields, channels, baselines and times) can be given to define the size of the MSs. Furthermore, a tile (chunk) shape can be given to optimise non-sequential data retrieval.
- readms to read (part of) an MS using a given access pattern.

When multiple MSs are created, the commonly used spectral data distribution is used, thus each MS contains a subset of the frequency channels.

Benchmark results

The main tests have been done on a LOFAR-like MeasurementSet with 1953 baselines, 384 subbands of 64 channels and 4 polarisations. Writing an MS took about twice as long as writing a raw data file of the same size, even when tiling the MS data. Writing the similarly sized HDF5 file took about as long as writing the MS, also when chunking the data. Thus, in this respect, there is no performance difference between HDF5 and MS.

Reading the data in a sequential way shows the same performance as writingfor MS and HDF5. However, for non-sequential access HDF5 is much slower when small data chunks have to be read.

To measure and compare reliably, the scaling tests have been performed with a various number of time steps. These tests show linear scaling for both MS and HDF5.

A recent addition to the underlying Casacore infrastructure is the support of the ADIOS2 framework making it possible to do parallel IO. These benchmark tests will be performed in the future. Preliminary results were presented in [15], more information on the package can be found in [16].

8. CTA-DAS Comparison of data formats for CTA data. (UCM - Quasar S.R.)

CTA, the Cherenkov Telescope Array will be the first open Cherenkov observatory. Different prototypes are being developed for its software and data format. The goal of the project was to explore data format for the storage of low level data. This work, if successful, could be used by other Cherenkov observatories of the same type.

Data Level 0 (DL0) refers to the data obtained during the acquisition process, either by the telescopes (Hardware HW) or by simulations (Software SW). In the data flow of CTA, DL0 is the first data level that is permanently stored, and it serves as the main input for a pipeline that creates Data Level 1 (DL1) products.

The DL0 products consist of files with lists of thousands of events (coordinates, time, energy, etc.) captured by the telescopes. There are several ways to store this data, but its complexity and size make it difficult to manage in an efficient way. For that, the testing of different file formats and technologies has to be done in order to determinate which one is the most suitable for the task.

Due to its characteristics, the file format chosen for testing the CTA-DAS was the Hierarchical Data Format 5 (HDF5). HDF5 is a file format for extremely large and complex data collections, widely used in the scientific community as well as the general industry. The goal is to develop a software capable of creating DL0 files and compare its performance against other formats been under study, or development, within the CTA collaboration. The software also needs to be able to convert the current CTA format (CTA LST R1) into the HDF5 format proposed.

HPY is a Python 3.6 library developed by Quasar Science Resources, S.L. that handles the transformation of CTA LST R1 Data into the HDF5 format. The library also provides a module to read general files in ZFITS format, which is used to perform the conversion.

Benchmarks

In order to test the viability of this HDF5 format within the CTA project, performance benchmarks were developed. Testing how good the performance of the HDF5 is, was one of the main goals of this software. Because of this, the library allows the user to choose between different modes, different compression methods, and HDF5 libraries.

The HDF5 libraries used were h5py and pytables. Tests show, that out of these two, the one giving the most promising results was h5py, both in speed as well as in size of the final files.

Two modes were developed within the HPY library. The different modes allow creating two different HDF5 formats. The first mode creates HDF5 files grouping the event data by HDF5

groups. The second mode groups the events by tables. Grouping the events by groups is faster but leaves the file with thousands of groups, which is not very convenient to read. Grouping them by tables, although slower, allows easing the way the data is read.

Also, it is worth mentioning that at the beginning of the development, the library used to read the LST R1 ZFITS was *astropy*, which does not decode the ZFITS. This means that the files created using this library to read the input files provides coded HDF5 that need to be decoded afterward by the user. This was rapidly changed and the software now uses the protozfits library provided by the UCM to solve this problem.

Of all the compressing methods used by HPY, the one with the best compression rate is gzip, but, as expected, the compression process slows the creation of the HDF5 significantly.

The benchmarks conclude that the HDF5 format has its pros and cons and depending on how the files are created the performance can be affected significantly, but overall is a good option to manage the CTA DL0. Full detail of the benchmarks realized can be found in [17].

9. IDG - Image Domain Gridding (ASTRON)

For wide field instruments like LOFAR and in the future SKA-LOW it is essential to take direction dependent gain variations into account when constructing an image from observed correlation data. Especially for effects varying on short time scales, such as the ionospheric effects, this is computationally expensive.

The Image Domain Gridding (IDG) algorithm can apply these corrections efficiently and was designed to be well suited for implementation on Graphical Processing Units (GPUs).

The implementation for various architectures is available at https://gitlab.com/astron-idg/idg. The derivation of the algorithm, and its numerical and computational performance has been published [1]. A recent analysis [2] on gridding accuracy specifically for the Epoch of Reionization (EoR) concludes that IDG is the best-suited gridding algorithm for this science case, including future SKA EoR observations.

Another time consuming step in wide field imaging is estimating the direction dependent corrections, or direction dependent calibration. Initial tests using IDG inside the calibration loop are promising. Development of this method is ongoing.

The WSClean imager, available at https://sourceforge.net/projects/wsclean now supports an IDG gridding mode.

Currently this mode supports corrections for

- LOFAR beam

- Murchison Widefield Array (MWA) beam
- User supplied corrections in the form of FITS images

WSClean, including IDG functionality, is used in the prefactor pipeline, available at https://github.com/lofar-astron/prefactor.

The LOFAR Radio Observatory will make prefactor processing of observations available to LOFAR users [18, 19, 20]

10. Conclusions

This deliverable shows a summary of some of the work developed inside the D-GEX, package 3.2 of OBELICS. Although it is not a complete review, it allows the reader either a scientist or a policy maker to get an overview of the work developed. Originally conceived as a set of benchmarks on existing software, since the resources committed by OBELICS have allowed the groups to go further and help to develop innovative solutions, we have considered it convenient to describe them here, together with the benchmarks performed.

The report covers contributions from several ESFRI projects: CTA, KM3NeT, SKA. The software is available in open repositories, notably the ASTERICS repository and will be of use to many other researchers. In many cases, as the BASC source algorithm, the contribution to DL3 format and convertors, or the introduction of fluorescence in CORSIKA, the work performed was published in high impact peer reviewed scientific journals.

In addition, a community has been created whose impact in astronomy will be considerable in the year to come.

11. References

[1] Aubert, P., Vuillaume, T., Maurin, G. et al. Comput Softw Big Sci (2018) 2: 6. https://doi.org/10.1007/s41781-018-0010-3.

[2] Bayesian source discrimination in radio interferometry. Hague, P. R.; Ye, H.; Nikolic, B.; Gull, S. F. Monthly Notices of the Royal Astronomical Society, Volume 484, Issue 1, p.574-581.

The BASC package, developed by P. Hague is available at https://github.com/petehague/BASC

[3] The OMGSIM package, developed by C.Hugon (CPPM) and V. Kulikovskiy - INFN Genova, is available at https://github.com/vkulikovskiy/OMGsim

[4] The ParamNu package, developed by L. Quinn and D. Zaborov (CPPM), is available at https://git.km3net.de/lquinn/param[18nu-public.

[5] The MAGCI to DL3 convertor, developed by T.Hassan (IFAE) , C. Nigro (DESY) and L. Jouvin(IFAE) is available at (https://gitlab.com/magic_dl3/magic_dl3

[6] D. Heck, J. Knapp, J.N. Capdevielle, G. Schatz, T. Thouw, CORSIKA: a Monte Carlo code to simulate extensive air showers, Forschungszentrum Karlsruhe FZKA 6019 (1998) 1–90.

[7] S. Martinez, F. Arqueros, V. Fonseca, Monte Carlo simulation of the HEGRA cosmic ray detector performance, Nucl. Instrum. Methods Phys. Res., Sect. A 357 (2–3) (1995) 567–579.

[8] K. Bernlöhr, Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_telarray, Astropart. Phys. 30 (3) (2008) 149–158.

[9] B.S. Acharya, M. Actis, T. Aghajani, Introducing the CTA concept, Astropart. Phys. 43 (2013) 3–18.

[10] D. Morcuende, J. Rosado, J.L. Contreras, F. Arqueros, Relevance of the fluorescence radiation in VHE gamma-ray observations with the Cherenkov technique, Astropart. Phys. 107 (2019) 26-34.

[11] M. Reininghaus, R. Ulrich, CORSIKA 8 - Towards a modern framework for the simulation of extensive air showers, Proceedings of Ultra High Energy Cosmic Rays 2018, Available at https://arxiv.org/abs/1902.02822.

[12] K. Bernlöhr, Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_telarray, Astropart. Phys. 30 (3) (2008) 149–158.

[13] D. Morcuende, J.L. Contreras, J. Rosado, A Monte Carlo study of the relevance of fluorescence radiation in VHE gamma-ray observations with Cherenkov telescopes, Proc. Sci. (ICRC2017) 839.

[14] M. Tluczykont, D. Hampf, D. Horns, The HiSCORE concept for gamma-ray and cosmic-ray astrophysics beyond 10 TeV, Astropart. Phys. 56 (2014) 42–53.

[15] G. van Diepen, “Casacore Table Data System and its use in the MeasurementSet”, Astronomy and Computing, September 2015

[16] https://www.hdfgroup.org https://csmd.ornl.gov/software/adios2

[17] Reference for HPY. CTA Data Management Technical Design Report

[18] Sebastiaan van der Tol, Bram Veenboer and André R. Offringa, Image Domain Gridding: a fast method for convolutional resampling of visibilities, Astronomy & Astrophysics, Volume 616, A27 (2018), https://doi.org/10.1051/0004-6361/201832858

[19] A. R. Offringa et al., Precision requirements for interferometric gridding in 21-cm power spectrum analysis, (submitted to A&A, April 2019)

[20] de Gasperin et al., Systematic effects in LOFAR data: A unified calibration strategy, Astronomy & Astrophysics, Volume 622, id.A5, 18 pp, https://doi.org/10.1051/0004-6361/201833867