ANNA BARNACKA
LESZEK BOGACZ
MIRA GRUDZIŃSKA
MATEUSZ JANIAK
NUKRI KOMIN
GIOVANNI LAMANNA
RAFAL MODERSKI

MONTE CARLO SIMULATIONS FOR THE CHERENKOV TELESCOPE ARRAY OBSERVATORY USING PL-GRID E-INFRASTRUCTURE

Abstract   *The paper presents Monte Carlo simulations carried out during the preparatory phase of the Cherenkov Telescope Array project. The aim of the project is to build the next generation observatory of very high energy gamma rays. During the preparatory phase there is a need to optimize and verify design concepts for various elements of the array. In this paper we describe the main components of the software being used for that purpose, their functions and requirements. Preliminary results of the optimization of the small telescope – one of the several kinds intended for the array, are presented.*

1. Introduction

Measuring cosmic radiation at very high energies (VHE), above 100 GeV, requires more than a simple telescope with a CCD camera. The latest generation of gamma-ray instruments like H.E.S.S. [1], MAGIC [2] and VERITAS [3] consists of multiple telescopes with analogue cameras based on photomultipliers. This is the only way to observe Cherenkov light from extensive air showers that are produced when a high energy particle enters the Earth's atmosphere. Those experiments have delivered spectacular astrophysical results. Now, based on the experience of current instruments the community of more than 100 institutes from 25 countries is going to build a new generation ground-based gamma-ray observatory – the Cherenkov Telescope Array (CTA)[4]. The array will consist of tens of telescopes and will have an order of magnitude better sensitivity as compared to currently operated instruments.

Because of the complexity of the project a proper preparatory phase is required before the actual construction of the facility. The CTA will operate at least three kinds of telescopes, each one equipped with a high sensitivity analogue or digital cameras based on standard or semiconductor photomultipliers. Cameras will detect even single Cherenkov photon, then the trigger system will determine if the event observed is scientifically interesting or not.

The overall performance of the system depends on many parameters such as the configuration of the telescopes, size of a single telescope (aperture, focal length, field of view), the parameters of the cameras (the size of a single pixel, the amount of pixels, the sampling) and many more. To determine the best configuration at optimal cost billions of numerical simulations are required. This means hundreds of TB of storage and many years of single CPU time. During the operational phase CTA will produce at least 3 PB of data per year. All these data will need to be analyzed, archived and shared with the community. Additional massive Monte Carlo simulations will be required to provide calibration for the scientific data. The high data rate of CTA together with the large computing power requirements for the Monte Carlo simulations demand dedicated computer resources which can be well handled through the grid approach. The EGI grid (`www.egi.eu`) [5] infrastructure and middleware for distributed computing, data storage and access are considered the most efficient solution for the CTA e-infrastructure.

In our previous paper [6] we have described the details of the numerical simulations performed during the, so called, Prod-1 phase. Here, in Sec. 2 we briefly remain the procedure of generation and the analysis of the Monte Carlo data. In Sec. 3 we present new results of the performace studies of the array of small size telescopes. We devote Sec. 4 to our conclusions and prospects for future research.

2. Monte Carlo simulations and analysis

The CTA project is currently in its preparatory phase. One of the main purposes of this phase is to find the layout of the array. The optimal parameters need to

be estimated both for a single telescope: mechanical details, such as construction geometry, mirror layout, and also camera details, electronics, etc. and for the whole array – telescopes layout and spacing. To study the performance of the array massive Monte Carlo simulations are performed. These simulations consist of several steps. First, the development of the extensive air showers (EASs) caused by high energy particles is calculated and the resulting information about Cherenkov photons capable of reaching the detector is stored in output files. In the second step the detector response to the Cherenkov photons is simulated, while in the final step the analysis of the obtained data is performed.

The extensive air showers are simulated with CORSIKA (COsmic Ray SImulations for Kascade) [7, 8] – a numerical code which uses Monte Carlo technique to calculate the interactions between the particles involved in the development of the shower. Two examples of simulated showers are presented in Fig. 1. An additional package takes care of the emission of the Cherenkov light – a faint, blueish light from relativistic particles traveling through the Earth's atmosphere.
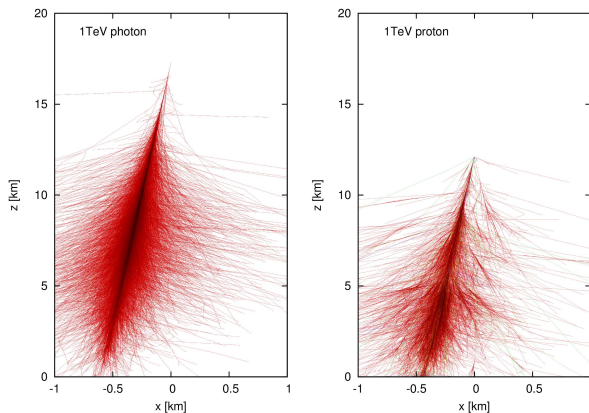

Figure 1. Two examples of simulated extended air showers developed in the Earth's atmosphere. Each line represents the path of an elementary particle created during the shower evolution. The left panel presents the shower generated by a single photon of energy 1 TeV, while the right panel shows the shower originating from a single proton of the same energy

CORSIKA CPU and memory requirements depend, to a great extent, on the type of primary particle and, of course, on the desired number of showers to be simulated. While the low energy ($\sim 100\,\mathrm{GeV}$) gamma photon simulations with several hundred thousand showers can take just a few days on a single CPU, for high energy events with energies of 100 TeV this time can grow up to over a year due to larger numbers of secondary particles and thus interactions to be computed. The CORSIKA output files can have sizes about several GB for a single telescope up to several hundred GBs for an entire array. Simulations for several values of gamma energies and several other particle types (such as protons, electrons, and muons) need to be performed.

The nature of Monte Carlo simulations requires at least a few thousands of events for each energy and each particle type in order to properly describe reality in a statistical sense.

The second performance estimation step is based on `sim_telarray` software which simulates all the telescope hardware response to Cherenkov photons [10]. The aim of this step is to get a final, raw output similar to the final output of a fully operational array. `Sim_telarray` provides a list of events detected by the camera of the telescopes. A graphical example of such an output is shown in Fig. 2.
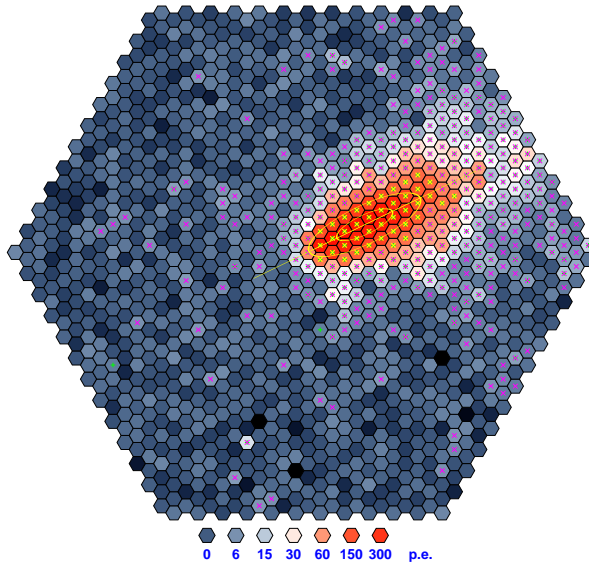
Figure 2. Example output from `sim_telarray` showing telescope camera triggered by a single incident gamma ray. All pixels show some signal due to the night sky background and electronic noise

After computing the telescope response to the air showers it is then necessary, in the final step, to translate the hardware response to measurable values. The analysis of the `sim_telarray` output is done using the `read_hess` software which attempts to reconstruct events seen by the telescope by applying various image analysis methods, image cleaning etc. The procedure must be repeated many times since usually we do not know in advance which parameters should be used for the analysis.

To sum up, the Monte Carlo simulations performed during optimization of the CTA array are a very complicated task and demand large resources, both CPU time and storage. The analysis usually requires more than 10^6 air showers to be simulated for each kind of particle and for several different energies. The process produces tens of TBs of data, and uses thousands of CPU hours. The `sim_telarray` analysis is less storage consuming, but since many hardware configurations need to be considered it also involves thousands of CPU hours. The output from `read_hess` occupies only tens

of MBs, but it must be produced for every `sim_telarray` job and thus is comparably CPU consuming. Since none of the software uses parallel computing the whole analysis procedure is also input/output intensive. The InfiniBand network connection thus becomes highly desirable. Such resources are difficult to collect in a single computer cluster and the grid approach seems to be more appropriate in the case of CTA simulations. To provide the necessary infrastructure the virtual organization (VO) `vo.cta.in2p3.fr` was created within the EGI grid. PL-Grid with its ZEUS computer cluster located in ACK CYFRONET is one of the largest computer centres supporting the virtual organization of the CTA.

3. Results

The first part (Prod-1 phase) of the massive Monte Carlo simulations for the CTA project started in 2008 using the resources available for `vo.cta.in2p3.fr`. During these simulations PL-Grid was the second largest resource provider for the VO with more than 100 TB of storage and over 2000 CPU cores. The main goal of the Prod-1 production was to simulate showers initiated by cosmic rays. These are mainly charged particles ($\sim 99.9\%$), like protons ($\sim 89\%$), α-particles ($\sim 10\%$), ionized nuclei of heavy elements ($\sim 1\%$), electron/positrons ($\sim 1\%$) and a small fraction of γ-rays. It was necessary to simulate a huge number of showers (e.g. $\sim 10^{10}$ protons) before estimating the remaining background and drawing conclusions about the performance of a particular configuration.

The CTA array will consist of at least three kinds of telescopes: large size (LST), medium size (MST), and small size telescopes (SST). During the Prod-1 phase the generic configuration of 275 telescopes of 5 different types has been considered (see Fig. 3). A candidate configuration being a subset of this (typically about 80 telescopes of total estimated cost of approximately 80 million EUR), marked by letters A-K were later chosen for detailed analysis.

The result of the CORSIKA simulation, which is the information on the showers which are theoretically detectable by at least one telescope was stored. Until now (June 2011) 7.9×10^3 runs with 7.9×10^8 showers were done. From this amount 2.3×10^7 were triggered by at least one telescope. The amount of data currently available for further analysis exceeds 100 TB.

In addition to the Prod-1 simulations we have performed dedicated massive simulations to study various trigger strategies and the performance of arrays consisting of just SSTs. We have used 40 different SSTs to form two kind of arrays: E3 which is similar to CTA-E from Prod-1 phase, but with no medium and large telescopes and four additional telescopes in the middle of the array to fill the gap, and regular arrays with telescopes distributed on a regular grid. The array layout can be seen in Fig. 4.

For both arrays we simulated air showers from gamma-rays, proton and night-sky background. Those events were processed using the `sim_telarray` software. Since our task was to study different combinations of telescope parameters we could not use `sim_telarray` to directly read CORSIKA output since there was a risk that
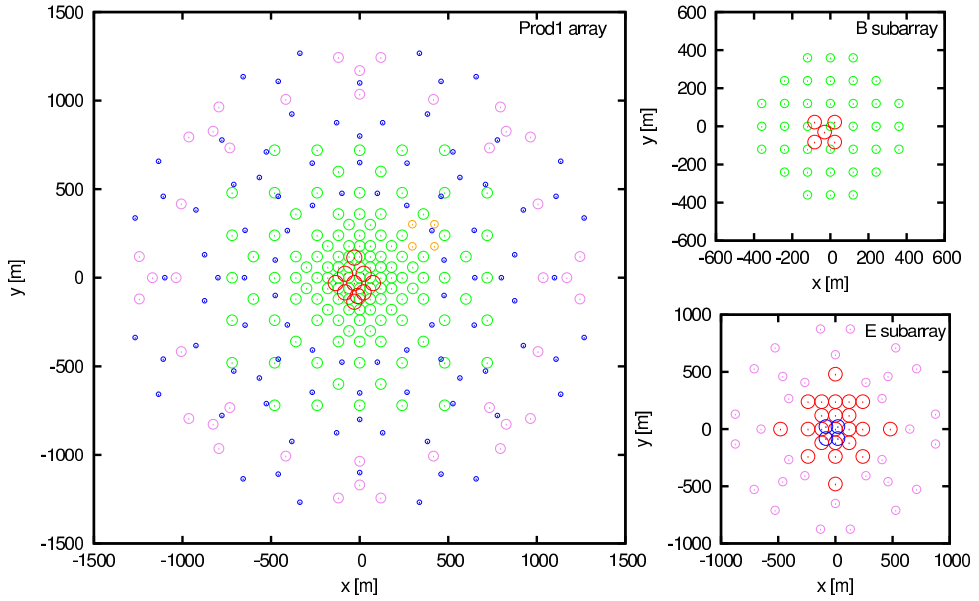
Figure 3. The generic configuration of 275 telescopes for Prod-1 phase and two example candidate configurations B and E. Taken from [9]
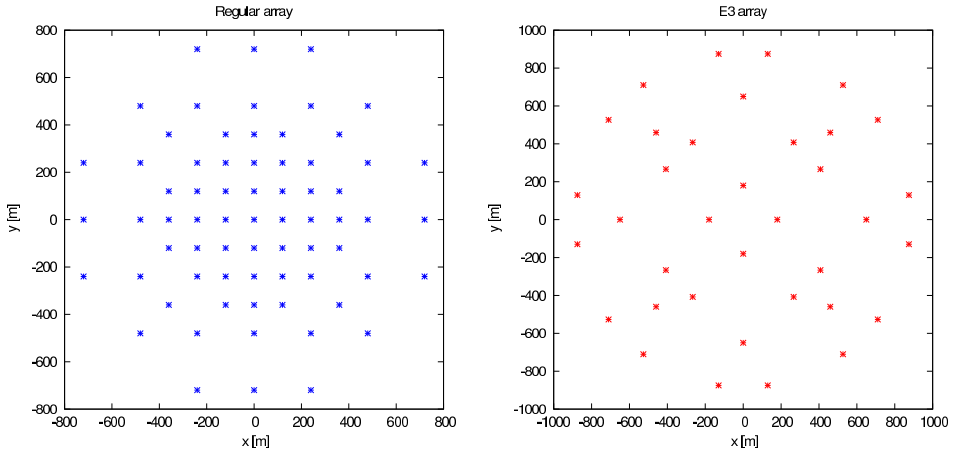

Figure 4. Two array layouts used to study the influence of a single telescope configuration on overall array performance. Each point represents a single small size telescope

this would lower the performance or even cause data loss. Instead, we first produced CORSIKA files and then used `sim_telarray` to analyzed them. For the array studies 11 energies were simulated (taken from 500 GeV to 100 TeV) and for single telescopes

30 energies were simulated (from 50 GeV to 100 TeV). All the data was analyzed using the `read hess` code to obtain the final results.

In order to analyze trigger strategies we calculated telescope effective areas and trigger efficiencies. Also, the night-sky background level was measured which is a crucial parameter for Cherenkov telescopes. The results are presented in Fig. 5 and Fig. 6.
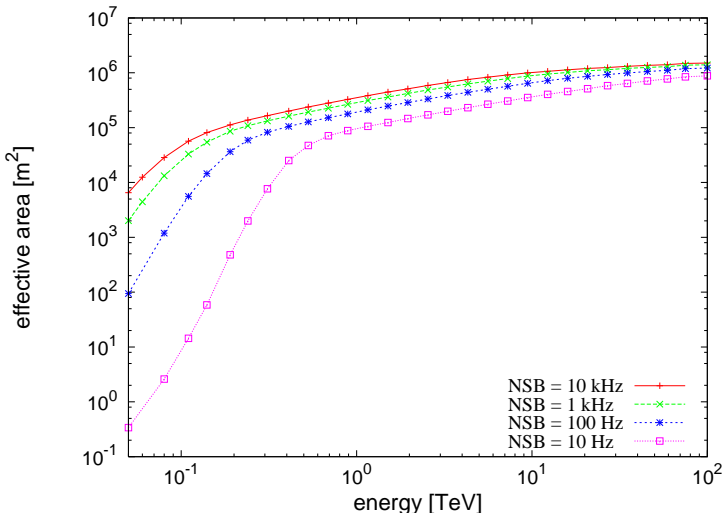
Figure 5. Effective area of the single SST as a function of energy for different incidental NSB levels

For a given night-sky background rate, we have calculated also a total data rate that the camera electronics has to manage in real operation. This is important information for designing electronics and read-out systems as well as the data acquisition system.

The study of the telescope array provides quite a challenge since the number of data one has to manage is significantly bigger than for a single telescope. Also, the CPU time is much longer because in the case of an array, not only single telescope triggers are taken into account, but there is also a central trigger system responsible for reducing data rates from muons and night-sky background. For an array simulation each telescope has to communicate with each other so the required CPU time is much bigger and does not scale linearly with a number of the telescopes. Also the analysis done with `read hess` is far more challenging. Since telescopes in an array are working together as a "single dish" each telescope output has to be verified with other telescopes to get a proper result. Full array simulations allow to calculate the sensitivity of the array and its angular resolution – two parameters that actually describe overall system performance.
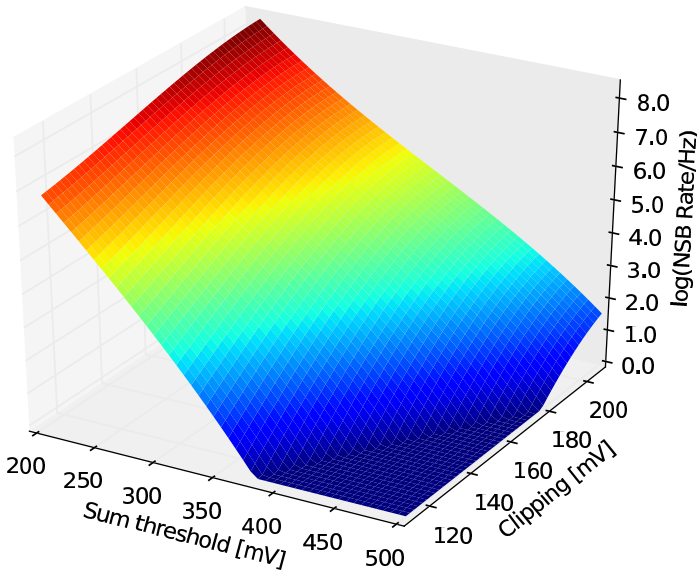
Figure 6. Surface plot of the NSB rate for different trigger parameters: sum threshold and clipping

The computational resources needed to perform the above-mentioned simulations were supplied by the PL-Grid project – mainly the ZEUS cluster at ACK CYFRONET in Cracow and a site CAMK located at the Nicolaus Copernicus Astronomical Center in Warsaw. The jobs were submitted through PBS queue system. The total estimated CPU time we used to simulate two arrays is estimated to be about one million hours (for CORSIKA, `sim_telarray` and `read_hess`) and about 5% of that value for single telescopes. The number of submitted jobs was about 4000, and the longest ones took more than 1000 hours of CPU time. The total amount of data is currently about 11 TB of which roughly 60 % is taken by larger regular array data.

4. Conclusion and future work

The analysis process was briefly described, and some of the preliminary results were presented. The amount of computer resources used (in both CPU time and storage space) even for such a basic analysis was noticeable. This gives the first view on how demanding is the task of complete array optimization, when the analysis of a large parameter space is needed. To draw realistic conclusions from the simulations one has to also develop the cost model. The overall budget of the project is known, and should be taken into account during optimization. The best strategy would be probably to search through parameter space only for configurations with the total cost equal the

budget. This would require realistic estimations of the price of individual components and solutions. We will investigate that subject during further study. Last, we will search for the optimal layout of the array depending on the number of telescopes, their configuration, type, and overall performance.

References

[1] Vasileiadis, G.: *The H.E.S.S experimental project.* [in:] Proceedings of the Fifth International Workshop on Ring Imaging Detectors, Nucl. Instr. and Methods in Phys. Research A, **553** (2005) 268–273

[2] Ferenc, D. and MAGIC Collaboration: *The MAGIC gamma-ray observatory.* Nucl. Instr. and Methods in Phys. Research A, **553** (2005) 274–281

[3] LeBohec, S. *et al. Deployment of the VERITAS observatory.* J. Phys.: Conf. Ser. **47** (2006) 232–237

[4] Glicenstein, J.-F. for the CTA Consortium: *The Cherenkov telescope array, an advanced facility for ground based gamma-ray astronomy.* [in:] Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors, Nucl. Instr. and Methods in Phys. Research A, **639** (2011) 46–49

[5] Ghiselli, A., and Mazzucato, M.: *Grid: From EGEE to EGI and from INFN-GRID to IGI.* Il Nuovo Cimento C **32** (2009) 233–241

[6] Barnacka, A. *et al.*: *PL-Grid e-Infrastructure for the Cherenkov Telescope Array Observatory.* [in:] Bubak, M., Szepieniec, T., and Wiatr, K., eds, *Building a National Distributed e-Infrastructure – PL-Grid*, Lecture Notes [in:] Computer Science vol. 7136, Springer 2012

[7] Capdevielee, J. N. *et al.*: *Extensive air shower simulations with the CORSIKA program.* [in:] L. Jones, ed., *Very High Energy Cosmic-Ray Interactions*, AIP Conference Series, **276** (1993) 545–553

[8] Heck, D., & The KASCADE Collaboration: *Hadronic Interaction Models and the Air Shower Simulation Program CORSIKA.* [in:] Proceedings of the 27th International Cosmic Ray Conference, International Cosmic Ray Conference, **1** (2001) 233

[9] Actis, M. *et al.*: *Design Concepts for the Cherenkov Telescope Array: an advanced facility for ground-based high-energy gamma-ray astronomy.* Experimantal Astronomy, **32** (2010) 193–316

[10] Bernloehr, K.: *CTA simulations with CORSIKA/sim_telarray,* [in:] F. A. Aharonian, W. Hofmann, & F. Rieger, eds, *HIGH ENERGY GAMMA-RAY ASTRONOMY: Proceedings of the 4th International Meeting on High Energy Gamma-Ray Astronomy.* AIP Conference Series, **1085** (2008) 874–877

Affiliations

Anna Barnacka
   Nicolaus Copernicus Astronomical Center, Warsaw, Poland and DSM/IRFU/SPP,
   CEA/Saclay, Gif-sur-Yvette, France

Leszek Bogacz
   Department of Information Technologies, Faculty of Physics,Astronomy and Applied
   Computer Science, Jagellonian University, Krakow, Poland

Mira Grudzińska
   Astronomical Observatory, Warsaw University, Warsaw, Poland

Mateusz Janiak
   Nicolaus Copernicus Astronomical Center, Warsaw, Poland

Nukri Komin
   LAPP, Universitas de Savoie et IN2P3/CNRS, Annecy, France

Giovanni Lamanna
   LAPP, Universitas de Savoie et IN2P3/CNRS, Annecy, France

Rafal Moderski
   Nicolaus Copernicus Astronomical Center, Warsaw, Poland