











# Investigation of the Capability of Restoring Information on the Primary Particle from Cherenkov Light Generated by Extensive Air Showers Using the Lomonosov-2 Supercomputer

*Elena A. Bonvech*<sup>1</sup> , *Clemence G. Azra*<sup>1,2</sup> , *Olga V. Cherkesova*<sup>1,3</sup> ,  
*Dmitriy V. Chernov*<sup>1</sup> , *Elena L. Entina*<sup>1</sup>, *Vladimir I. Galkin*<sup>1,2</sup> ,  
*Vladimir A. Ivanov*<sup>1,2</sup>, *Timofey A. Kolodkin*<sup>1,2</sup> ,  
*Natalia O. Ovcharenko*<sup>1,2</sup> , *Dmitriy A. Podgrudkov*<sup>1,2</sup> ,  
*Tatiana M. Roganova*<sup>1</sup> , *Maxim D. Ziva*<sup>1,4</sup> 

© The Authors 2024. This paper is published with open access at SuperFri.org

The new SPHERE-3 detector is under development. Its main objectives are the primary cosmic ray spectrum and chemical composition studies in the 1–1000 PeV energy range. The detector will register both reflected and direct Cherenkov light from extensive air showers. The goal of the new approach is high precision of event-by-event estimation of the primary particle parameters, especially its mass. The reflected Cherenkov light registration technique used in earlier experiments has good energy sensitivity and some mass estimation capability. Addition of direct Cherenkov light registration will allow to further advance the detector mass sensitivity. Several approaches to direct Cherenkov light registration are considered: by the main detector camera and by a dedicated direct light detector. First tests of the proposed methods are presented both for reflected and direct Cherenkov light. The detector design is tested on a large database of simulated showers. The simulation pipeline and related challenges to it are described. Also, progress in parallelization of the CORSIKA code for Cherenkov light simulations is presented.

*Keywords:* Cherenkov light, primary cosmic rays, supercomputer Lomonosov-2, extensive air showers, air-borne telescope.

## Introduction

Chemical composition of primary cosmic rays is crucial for the understanding of their sources, acceleration and transport processes. However, despite decades of studies, the precision of available data, especially at high energies, is somewhat insufficient.

Primary cosmic rays with energies above  $10^{15}$  eV are studied via indirect methods using large ground-based detector arrays using Earth's atmosphere as a natural calorimeter. Upon entry into the atmosphere a primary cosmic ray particle starts to interact with air initiating a cascade of secondary particles: an extensive air shower (EAS). EAS development in the atmosphere is accompanied by different electromagnetic phenomena, including Cherenkov light (CL) emission. In 1974 A.E. Chudakov proposed to register reflected CL from EAS [10]. In line of his idea we developed a series of SPHERE detectors [4]. The SPHERE-2 telescope operated in 2011–2013 and proved the validity and applicability of this method. Analysis of the results of the SPHERE-2 experiment gave an energy spectrum and chemical composition of primary cosmic rays in the 5–500 PeV energy range [5].

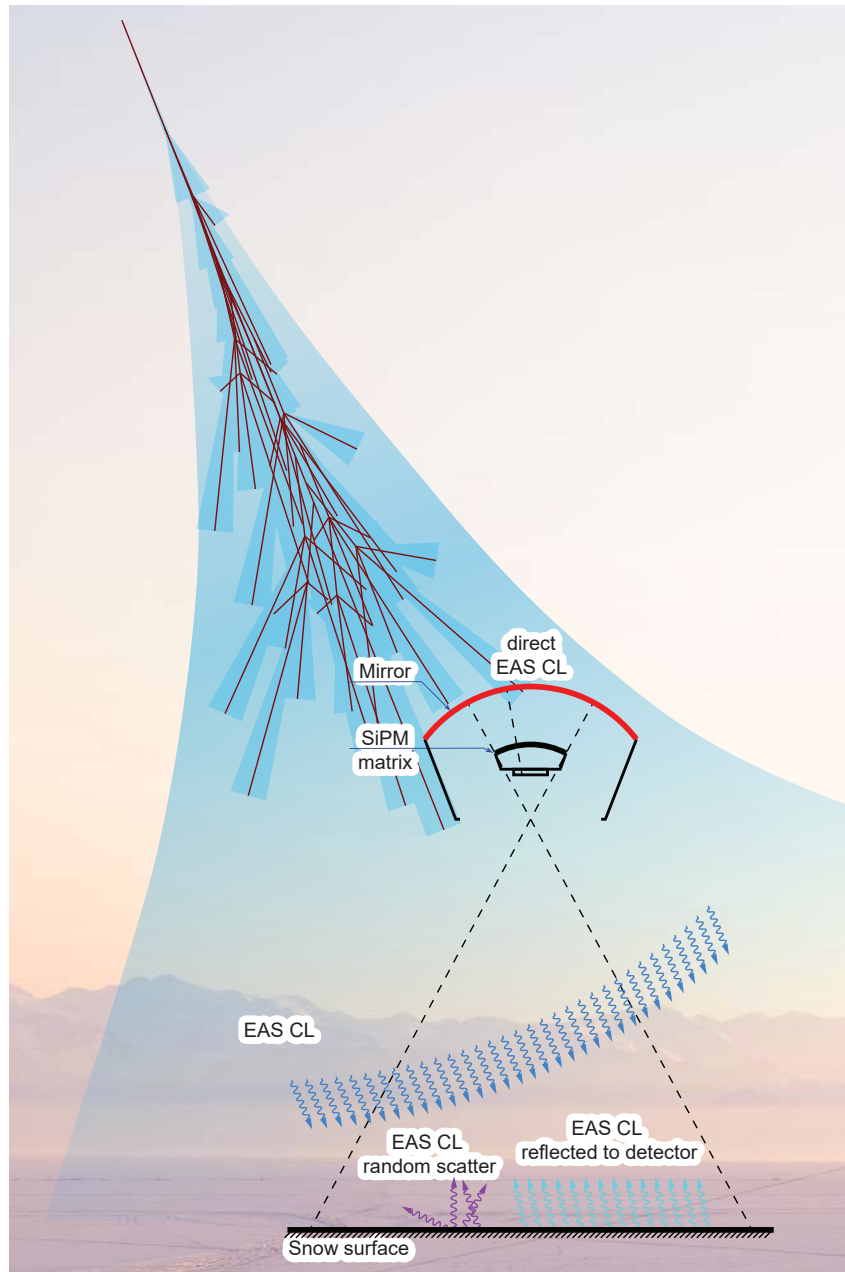
Currently, a new SPHERE-3 air-borne telescope aimed at reflected EAS CL registration is being designed. An unmanned aerial vehicle (UAV) is intended as a carrier for the detector. The proposed experimental setup is presented in Fig. 1.

<sup>1</sup>Skobeltsyn Institute for Nuclear Physics, Lomonosov Moscow State University, Moscow, Russian Federation

<sup>2</sup>Lomonosov Moscow State University, Physics Department, Moscow, Russian Federation

<sup>3</sup>Lomonosov Moscow State University, Department of Cosmic Research, Moscow, Russian Federation

<sup>4</sup>Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Moscow, Russian Federation



**Figure 1.** SPHERE-3 experiment scheme

The article is organized as follows. Section 1 is devoted to the SPHERE-3 detector description. The registration technique advantages, challenges and limitations are discussed. In Section 2 we present the simulations pipeline that allowed us to obtain data described and analysed later in the text. Section 3 reports on current progress in creation of parallel CORSIKA version that works with CL. In Section 4 we present first approaches for arrival direction and shower axis location estimation methods using reflected and direct EAS CL (independently from each other at this stage). Section 5 is devoted to current progress in our approaches to EAS primary particle energy and mass estimation, again independently, based on direct CL data and indirect CL data. Conclusion summarizes presented results and points directions for further work.

## 1. SPHERE-3 Telescope

The SPHERE-3 detector shares the same optical scheme with its predecessors: a Schmidt camera. The main mirror of the detector will be approximately 2.2 m in diameter and most probably segmented (there are some limitations in the different coating technologies). A corrector plate, absent in the previous SPHERE detectors, this time will be added, since the detector's light sensitive camera is planned to have a much higher resolution.

Substitution of the previous detectors photo-multiplier tubes by silicone photo-multipliers (SiPM) will allow to gain higher resolution and sensitivity of the detector camera, with lower mass and power consumption. The camera will consist of 379 optical modules with 7 SiPM each, 2653 SiPM in total. Higher resolution will give better precision for the reflected EAS CL distribution registration. Lower mass of the sensitive camera will allow a larger detector to be lifted on the UAV, besides, this less power consumption will allow more measuring channels, since each measuring channel with its electronics consumes power, as well as higher overall autonomy of the detector.

The planned flight altitudes are from 500 to over 2000 m. However, the exact measurement program will be corrected after more studies on the best flight parameters are done, since data registered at each altitude can yield information on a different energy region. The weather is also a factor to be accounted for.

### 1.1. Direct Cherenkov Light: EAS Primary Particle Parameters Estimation

Since earlier in the SPHERE project only the reflected EAS CL was registered some comments on why it was decided to include direct CL and what is to be achieved from analyzing its properties are needed.

All present day EAS detectors register only direct CL (if they register CL at all). SPHERE experiments were an exception, thus requiring us to develop new methods and approaches for data analysis. For direct EAS CL registration with ground-based detectors there exists an established set of methods with known capabilities and virtually no possibility to add some radically new approach. The primary particle energy is usually determined using some form of total CL photon count or by using photon density measures at some distance from the shower axis (150–200 m) [8, 16], where fluctuations are considered to be relatively small. The EAS arrival direction is reconstructed using delays between CL registration by several stations of the detector array or, alternatively, CL angular distribution can be studied at high resolution [6]. In this case, the maximum (or center of mass) of this distribution will indicate the shower arrival direction. However, this approach exhibits a systematic shift depending on the distance to the shower axis, shower zenith angle, primary particle energy and particle mass. But, the most difficult problem is the primary particle mass estimation. The generally used approach, first, requires to estimate the depth of the shower maximum  $X_{max}$  derived from the steepness of the lateral distribution function (LDF) for CL photons or from the CL pulse width at significant distances (above 300 m) from the shower axis. Then, using  $X_{max}$  and basing on the average model distributions with respect to primary particle energy and shower arrival direction the primary particle mass is defined. The pros and cons of this approach are analyzed in detail in earlier publications [11, 12]. Our current approach estimates the primary particle mass from the parameters of the CL distributions (spacial [11, 13] and angular [9]), making the resulting

criteria more integral in a sense that they utilize the whole distribution rather than its small part, with a low dependence on the used interaction model as a bonus.

The idea to introduce direct CL registration to the SPHERE-3 telescope came from two major sources: its accidental registration by the SPHERE-2 telescope [7] and transition from a balloon to an UAV as a carrier. *A priori* the idea does not look promising, since the limitations of the UAV platform are quite severe. However, first estimations show that even extremely small (compared to imaging air Cherenkov telescopes) direct CL detectors allow to obtain independent estimations of primary particle parameters, which is, arrival direction and mass. Especially interesting becomes the possibility of simultaneous registration of both direct and reflected CL from the same EAS allowing for two independent estimations of the shower primary parameters.

A direct CL detector alone cannot provide data for LDF analysis, only angular and temporal information is available for registration. But, the analysis of temporal data on EAS CL requires the detector to have nanosecond resolution and precise external knowledge on the shower axis location. And even fulfilment of these requirement does not guarantee any viable precision of primary mass reconstruction. On the other hand, it is known, that the CL angular distribution does hold information on the EAS longitudinal development [6], what makes it sensitive to the primary particle mass. As it will be shown below, even basic parameters of the angular distribution are sufficient for this task and allow to estimate the EAS arrival direction, but for best results they should be combined with reflected CL data. This brings us back to the idea of simultaneous registration of direct and reflected EAS CL. The SPHERE-3 detector construction and measurements strategy should be optimized so that most of the registered showers will contain data on both direct and reflected CL.

## 1.2. Limitations on Detector Location Relative to the Shower for Direct Cherenkov Light Registration

From the telescope target energy range of 1–1000 PeV and reasonable mass and size of the direct CL detector (a kilogram in mass and around 100 cm<sup>2</sup> area) come the limitations on shower core distance. At large distances, there will be not enough photons to reliably register and then reliably reconstruct the relevant distribution parameters. For the lower end of the target energy range (about 1–3 PeV) the distance to the shower core is estimated to be 100–200 m. Arrival direction estimates are possible in a broader range, but mass estimation at lower distances requires higher detector resolution as the shower angular images become smaller and axially symmetrical, while at larger distances the detector area should be bigger in order to collect enough photons.

## 1.3. Limitations of Detector Altitude for Direct Cherenkov Light Registration

The direct CL flux at a fixed distance from the shower core depends on the altitude: roughly weakening twice going from 500 to 2000 m. In the same manner, the size of the CL image shrinks with altitude. This leads to a higher energy threshold and higher required angular resolution of the detector to provide data for reliable primary mass estimation. Calculations show that the altitudes between 500 and 1500 m are the most favourable for our detector design. The expected fraction of EAS with both direct and reflected CL data will be around 0.3.

The properties of the angular distributions from direct CL obtained using the CORSIKA code, as well as the characteristics of images from a toy-model of a direct CL detector consisting

of a lens and a large area position-sensitive sensor were studied. The task was to separate EAS by their primary mass and estimate their arrival direction. Each of the tasks was solved separately for the angular distributions and for detector images. The careful analysis of model angular distributions allows to set the upper limits on the accuracy of parameter estimates in each of the two tasks, which will then be needed for the detector design. Solving these problems for a toy-model detector brings the accuracy closer to real values.

## 2. Calculations Pipeline

The design and optimization of primary cosmic rays optical detectors is a computationally heavy and complex procedure. This task includes not only optimization of the optical part (a telescope) that needs to meet certain criteria such as area, available materials and some specific needs that are significantly different from other optical instruments, but also includes simulation of EAS development, registration process, electronics response and data analysis procedure. And only in the end the final result can be evaluated to some extent. For common optics, there are parameters of the image that are crucial for later analysis. In case of EAS studies these parameters are relatively unknown. It is always a decision between larger entry window and larger field of view, between better resolution and limitations of light sensors, between more pixels and limitations on how much data can be digitized, stored and later processed. All this makes it virtually impossible to divide the optimization task into separate stages.

Thus, the detector optimization loop, therefore, includes a full Monte Carlo simulation of the EAS development process. The number of particles traced in the simulation is proportional to the primary particle energy. A single shower from a very high energy particle may take weeks to simulate (however, this is rarely done for obvious reasons). But, even with relatively low energies detector optimization requires to account variations in EAS development conditions (the atmosphere is not constant and weather changes affect the simulation outcome) and different high energy hadron interaction models. In a perfect situation, an optimized detector should not be sensitive to the model change, however, this is impossible to achieve by construction changes only. A combination of detector construction and data analysis procedures may solve such a task. As it was shown for the SPHERE-2 reflected CL data, there is a way to lower the influence of the nuclear interactions model uncertainty for this detector by careful data processing [17].

On top of a long optimization loop that includes EAS development, light collection simulation, data handling and evaluation of the obtained results (which usually are based on statistical methods) there is another simulation step in the middle that introduces additional uncertainties: the light registration process itself. The expected number of photons reaching the detector in the end is relatively small, the light registration process in the sensor should also be a full Monte Carlo simulation, since fluctuations are high. These simulations should account for the sensitive element operation. Both PMTs from the earlier experiments and SiPMs in the current detector rely on cascade amplification from a single photoelectron to the final anode current pulse. Beside this, there are side-effects of amplification, digitization, background and various sorts of interferences (however, the last are easier to exclude from the optimization process and try to avoid or to compensate *post factum* in real data).

The electronics design for the detector is also not set in stone and allows for parameter selection such as amplification and dynamic range, digitization frequency and event record length (and how deep a buffer can be). These decisions may affect the trigger system logic and detector energy thresholds. The overall detector optimization procedure may even include a task to find

the minimal viable parameters of the electronic systems to reach the overall detector goals, which are the primary cosmic rays energy spectrum and mass composition studied with reasonable event statistics behind each data point.

This gargantuan problem even for only direct or only reflected CL requires a careful approach and tremendous computational resources unavailable outside the supercomputer framework. Trying to optimize for both simultaneously just adds complexity to the task. On the bright side, we can directly estimate the best possible results obtainable with the detector if we exclude the background and electronics from consideration. This will allow to check if it is possible to obtain certain results given the natural fluctuations in EAS development. Inclusion of the background and electronics will affect the results, but we will have a starting point in detector shape and key parameters to track. Also, the first approximation for data analysis can be obtained from this data.

Also, in the search for such analytical methods and approaches neural networks may come in handy. In the recent years, neural networks find more and more use in scientific research. However, their introduction to a new field goes with comparison to traditional methods and techniques, since they are quite sensitive to noise in the data (and even to the type of noise) while not providing transparent (for the scientific community who are yet not quite accustomed to them) indications and measures of success or failure. They also require careful data preprocessing and large statistics to train upon, what brings back the computational burden (plus the computational resources for neural network training).

To approach the solution of this complex problem we decided to unify and standardize approaches and procedures. The SPHERE-3 software combines several modelling and data processing stages into a single calculations pipeline. Since the main aim of the current stage of SPHERE project is the experiment and detector design optimization such approach allows smooth recalculation from any point (except, probably, the very first step as it is very time-consuming).



**Figure 2.** SPHERE calculation pipeline

The first step in the calculations was EAS simulation using the CORSIKA 7.5600 code [14] with two different models of high-energy hadron interactions QGSJET01 [15] and QGSJETII-04 [18]. At this step, a set of simulation parameters was chosen, that included energy, zenith angle, primary particle type, atmosphere model. Simulation results were saved for each simulated EAS as a separate spatio-temporal distribution on the snow and a set of angular distributions for different observation altitudes and shower core distances. These simulated EAS formed a data base that was used for all consecutive simulation steps.

Next, in order to boost the statistics and save time each simulated shower was “cloned” 100 times with random shifts in axis location relative to the detector to simulate photons reaching the telescope entry window. Photon arrival times are updated respectively to the travelled distances. This step allows to use more data from the shower on later steps and to study the precision of the registration system and reconstruction procedures.

The next step involves tracing individual photons inside the detector from the entry window to the registration surface of the light sensitive elements. This step was done using Geant4 [1–3]. Detector elements, due to their complexity, were modelled as independent parts, saved as STL

files and imported into Geant4 using CADMesh [20]. This approach simplifies the modelling process, keeping the geometry precise and allowing flexibility in detector geometry manipulations, since no code needs to be updated, only the geometry files. Geant4 in its run transports the photons inside the detector with account for all reflections and scattering until the photon is absorbed (on the light sensitive surface or on some not absolutely reflective surface) or leaves the detector back through the entry window. Photons that hit the sensitive surface are recorded with their arrival time and other properties.

Next comes the electronics response simulation (if needed). At this step, the list of photons at the detector SiPM matrix is padded (if necessary) with background photons and converted to the output of the data acquisition system (DAQ) readout. Since the detector is at the earliest stages of development, no real electronics exist. But, an early prototype of the electronic system was tested. A small scale SiPM matrix was tested as part of the Small Imaging Telescope in Tunka Valley [19] and a DAQ prototype was designed and tested in the lab. The results of these tests were used for approximation of the SPHERE-3 electronics properties (amplification, SiPM behaviour, amplification effects on SiPM output pulse profile etc.). Optionally, other electronics effects can be applied such as non-linearity, baseline fluctuations, noise, buffer shift etc. The output of this stage closely approximates the expected output of a real detector and can be used to test various approaches for trigger system design, data processing techniques, etc.

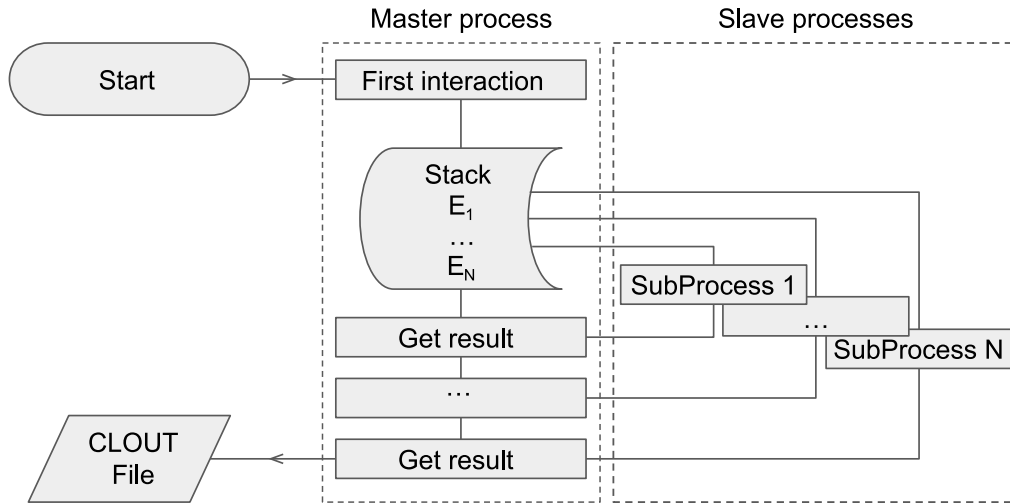
For convenience, all of the above mentioned steps were tied together into a unified pipeline using a Python application. This application allowed to select the desired simulation parameters (energy, zenith angle, nuclei, detector position, background etc.) and run the required simulation steps documenting the results in the process. Also, at any moment later, the pipeline may be expanded with any number of new steps allowing to include it naturally into the detector optimization loop.

### 3. Parallelized CORSIKA with Cherenkov Light Generation

The foundation of the simulations for the SPHERE project is a CORSIKA code (the standard for EAS simulations). While an option to compile a version of CORSIKA for parallel systems exists, it will only generate charged particles, and the authors of the CORSIKA code do not plan on making a parallel version with the CL option due to lack of resources. A single simulation of EAS with CL for a primary particle of 100 PeV will require at least 37 hours to complete, if all goes fast. If not (this is a Monte Carlo simulation through and through), it will take the simulation almost to the limit of 48 hours set for a job on most of the supercomputer systems. Thus, the required simulation precision makes it impossible to reach the top of the desired energy range (1–1000 PeV). In this situation, in order to perform simulations for even higher energies there are two options: to run simulations on local machines with no limitations or to create a in-house parallel version of CORSIKA using MPI libraries.

This first option severely limits the amount of showers that can be simulated at a reasonable time. Therefore, we started to develop a parallel version of the CORSIKA code. Since there was no aim at a generalized universal conversion of the code which would be compatible with other options of the software (and there are numerous options that provide required flexibility to the code) and there was no need to keep some of the CORSIKA base functionality (like charged particles distribution over the observation level, individual particle parameters, shower profile statistics etc.), only the required data on CL was to be kept and stored in form of angular-spatio-temporal distributions at certain levels. Along with the CORSIKA code native logic of keeping

a sort of a particles calculation queue (in the form of a particle stack) and physics that insure that the first interactions produce particles with highest energies, the parallelization scheme is rather simple (see Fig. 3).



**Figure 3.** Parallel CORSIKA operation scheme

Thus, a single job contains a separate task (Start block) that prepares every required bit of data for later processes (atmospheric parameters, cross-sections, particle tables and alike) and parameters of the event to simulate (primary particle, its start location, first interaction point, momentum etc.). All of this info is stored in a file. At the next stage, the master-process reads this file and starts tracking the first particle through the atmosphere where through the first interactions a stack of particles to be traced is formed. At the same time, a set of slave-processes is started that reads the same data file as the master-process and awaits for the data on the particles to trace from the master-process. The master-process distributes the stack of secondary high-energy particles and goes into listening mode. The slave-processes upon receiving their portions of the particles stack perform subshower simulations for those particles and send the collected data on CL photons to the master-process, which, in turn, aggregates the received data. When all slave-processes finish their tasks the master-process produces a file (CLOUT) and finishes the job.

At the moment the task of CORSIKA parallelization is at the stage of collecting the data required for subshower simulations to run and particle stack distribution algorithm design.

So far the base EAS CL distributions set from CORSIKA includes two separate runs (slightly different versions and distributions parameters) over 6 and 4 nuclei (first and second runs respectively) initiating showers with 3 different energies (however, more are to be simulated once the parallel CORSIKA version will be ready) in 4 different atmosphere models, 2 nuclear interaction models and 6 zenith angles. With each set of parameters 100 showers were modelled (total of 144 000 EAS) that alone took more that 140 000 node-hours to compute on the Lomonosov-2 supercomputer [21] (roughly an node-hour per EAS, but these results are for low energy EAS). At least about the same time will be needed to finish the CORSIKA simulation runs to get the full required set of energies.

But, even on this set of modelled EAS CL distributions the first results using basic approaches and previously tested methods were obtained.



## 4. Axis and Direction

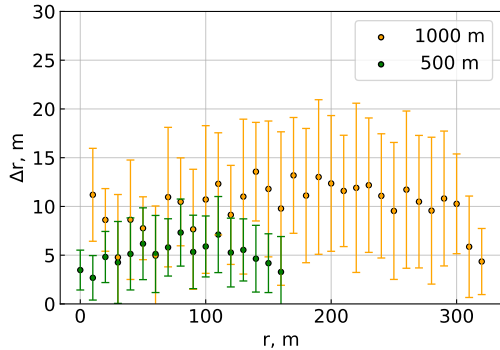
Axis location and shower arrival direction are the basic information that is reconstructed in EAS data analysis. The arrival direction can be estimated using both direct and reflected CL data.

### 4.1. Axis and Direction by Reflected Cherenkov Light

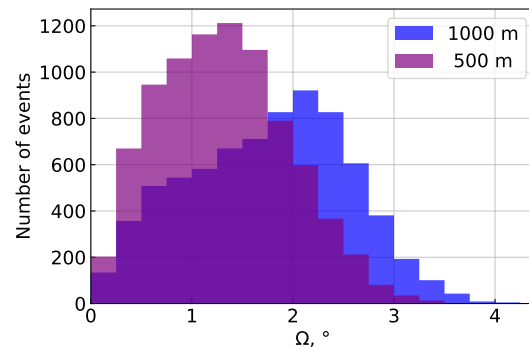
Reconstruction procedures were developed and tested on model events (without electronics response and background) from primary protons, nitrogen and iron nuclei with an energy of 10 PeV that arrived at  $15^\circ$ . Two detector flight altitudes were analysed: 500 and 1000 m.

Axis location was reconstructed from reflected CL data, at this stage, as a simple maximum search. During the first stage, an event from EAS is located within the recorded data frame (same as for the previous detector in the series, SPHERE-2) as a maximum in the time series of the total signal across all pixels. The assumption here is that while the amount of EAS CL photons is low relative to the expected background, they arrive in tight pulse and locally greatly exceed the average background values.

In the located event window individual pixels are analyzed, maximums in their time series are taken as data points (time and value). Across all values the pixel with the maximum value is identified. The data is then re-projected onto the snow and pixels are assigned coordinates with respect to the detector flight altitude. For higher precision weighted results from adjacent pixels are used. This is a simple procedure, but gives a relatively good precision – around 5 m for detector flight altitude of 500 m and around 10 m for 1000 m (see Fig. 4).



**Figure 4.** Axis location precision for different flight altitudes,  $\Delta r$  represents the distance between reconstructed and true axis locations



**Figure 5.** Arrival direction reconstruction precision using reflected CL data for different flight altitudes.  $\Omega$  is the angle between the reconstructed and true EAS arrival directions

Shower arrival direction was reconstructed from the same data pulses. Since the shower CL component is a thin (only few meters thick in its central part) slightly curved disk that effectively intersects the observation level, the light falls onto the ground as a thin line, travelling with a certain speed across the observation field. Time delays between the pulse’s appearance in each pixel (corrected by the optical path length to the detector) form a shower front in the  $(x, y, ct)$

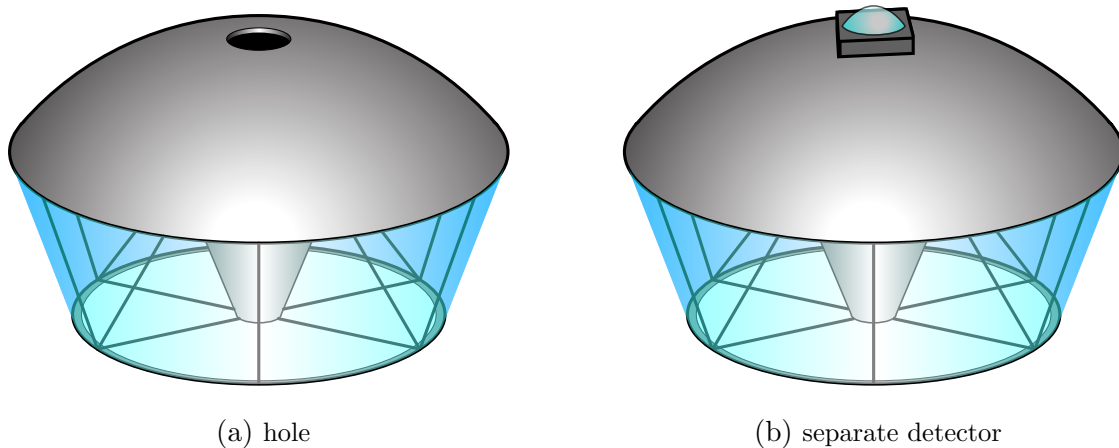
coordinates. This front was fitted by a parabolic function (in a EAS reference frame it forms a paraboloid of revolution around the shower axis):

$$t_i = a_0 + a_1 R(\phi, \theta) + a_2 R^2(\phi, \theta), \quad (1)$$

where  $R$  is the distance from the shower axis in a EAS reference frame,  $\phi$  and  $\theta$  are shower angles and  $a_i$  are the free coefficients, their dependencies on shower parameters are yet to be studied. The precision of this method of EAS arrival direction estimation is around  $1\text{--}2^\circ$  (see Fig. 5). Since the model EAS sample used in this initial study was small and all showers had the same zenith angle, the systematic uncertainties of this method are not yet studied.

## 4.2. Direct Cherenkov Light Arrival Direction

The proposed construction of the SPHERE-3 detector, specifically its UAV carrier, allows for two separate ways of direct CL registration: first, with the main SiPM mosaic through the hole (or a set of pinholes arranged into a coded aperture) in the main mirror (see Fig. 6a); second, with a separate compact detector (see Fig. 6b). Both have their pros and cons in capabilities, operation, procedures etc. This project is aimed, among other things, at selecting of the direct CL registration method.



**Figure 6.** SPHERE-3 versions with different approaches for direct CL registration

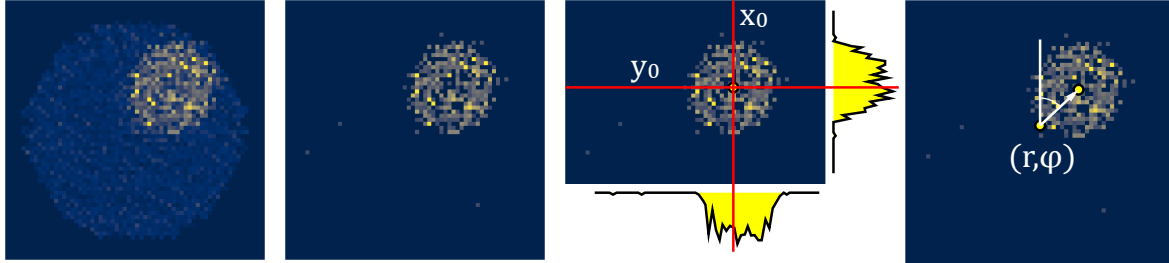
### 4.2.1. Main telescope mosaic approach

The difference in direct and reflected CL fluxes is so large that if a hole (or set of holes) configuration is used for direct light registration less than a 1% of the reflected aperture is needed, so the capabilities of the main mirror will not be affected.

As this is a work in progress, the simplest possible option was tested first – a large hole in the center of the mirror (the area of which is not involved in the reflected CL collection in any case). In the future, a more complex options may be evaluated for more data on other primary particle characteristics.

For this analysis a set of artificial events emulating EAS with a fixed brightness and varying angular distributions of zenith angles in a  $6\text{--}17^\circ$  range (azimuth angles were random) was used. These photons were transported through the detector and electronics response modelling. No

background was added, but random reflections and scattering of CL photons on various elements were accounted for. In case of an event (by the same method as was used for the reflected CL event location procedure described above) in each pixel the sum of a time series over a certain window was taken as the pixel value. Knowing the location of each pixel, this allows to construct an image (see Fig. 7) of the event.

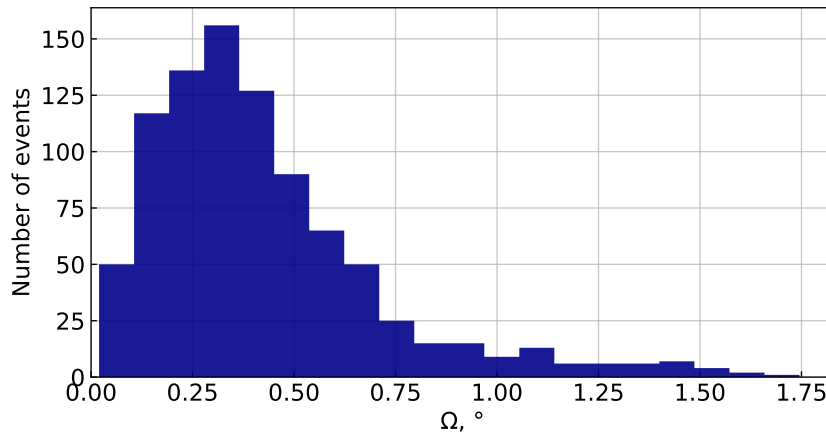


**Figure 7.** Direct CL images processing steps (from left to right): original image, cleanup, location of the center of the spot, feature determination

The images were then processed in the following steps:

- cleanup using the median value as a threshold, everything below is zeroed;
- the resulting spot's center was estimated as the center of mass;
- feature formation – distance between the center of the spot and the image center and arctangent of the spot's relative coordinates;
- a fully connected neural network with one hidden layer was used to reconstruct the true light arrival direction.

The resulting precision of the arrival direction determination was  $0.43^\circ \pm 0.29^\circ$  (see Fig. 8).



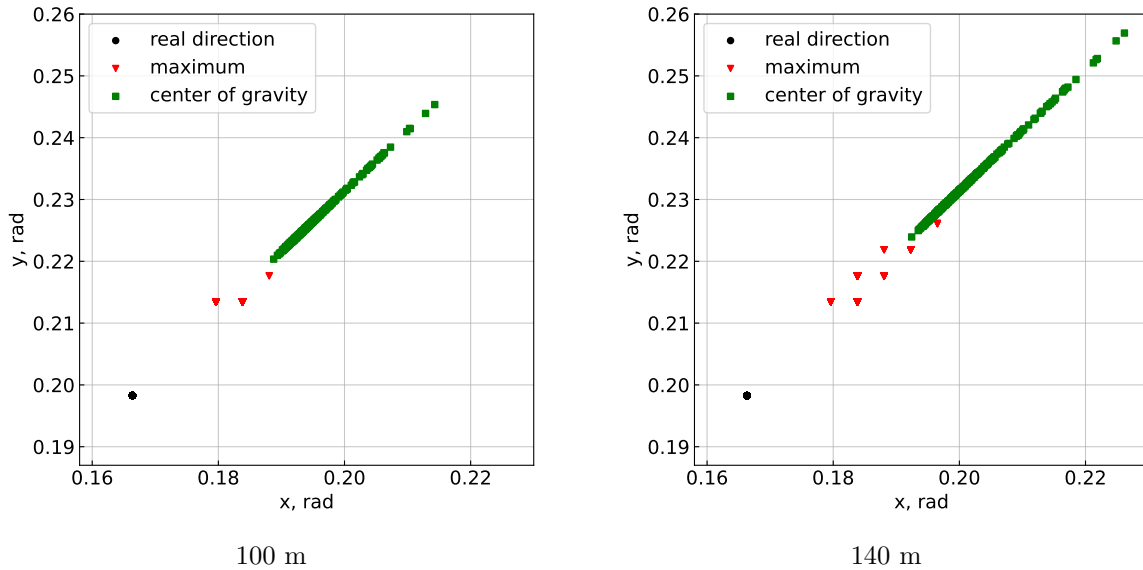
**Figure 8.** Arrival direction reconstruction precision using a neural network.  $\Omega$  is the angle between the reconstructed and true light arrival directions

#### 4.2.2. Using a separate detector for direct light

At the first stage for this approach, a pure EAS CL distribution was studied without any detector at a 100–200 m distance from the shower axis. In order to make the results comparable

and reasonable, photons that fell only on a small  $100 \text{ cm}^2$  area were used. This study allowed to evaluate the natural CL fluctuations and their effects on the expected results.

The photons that fell on this small area had an angular distribution – a subsample from the full EAS CL photon distribution at a given distance from the axis. This small distribution has a maximum and a weighted average (or center of mass), both shifted from the shower arrival direction. In Fig. 9, the results for two distances from a 10 PeV proton shower are shown. The samples were taken at 100 and 140 m from the shower axis at a fixed shower orientation. The shift of the CL spot's maximum (red dots, located in the vertices of the calculations grid), its center of mass (green dots) and the true shower arrival direction (black dot) are clearly seen.



**Figure 9.** Direct CL spot angular position (red and green, see text) relative to the true shower arrival direction at 100 m and 140 m distance from the shower axis

The spot position is shifted from the EAS arrival direction. However, the CL spot has an elliptical shape and the shift is along the spot long axis. The shift itself depends on the distance from the shower axis, EAS primary particle energy, type and arrival direction. But, this can be later accounted for. The precision of this method, even without corrections, is already high at around  $0.1\text{--}0.2^\circ$ , but, again, this result was obtained without interference from the background and real detector limitations.

A test with a simple one-lens detector (around 12 cm focal length,  $100 \text{ cm}^2$  aperture, high resolution detector) yielded virtually the same results in an analogous analysis without background.

## 5. EAS Energy and Mass Estimation

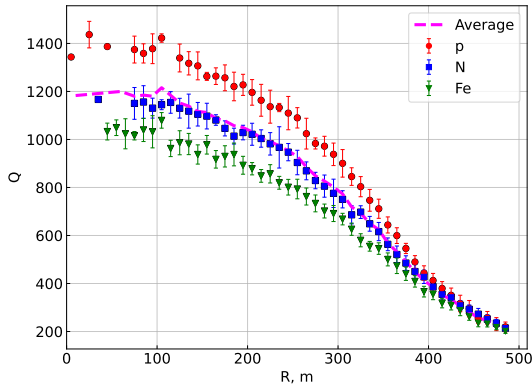
A more complex analysis is needed to estimate the EAS primary particle energy and mass, which are the main goals of this project. Some estimations were done, again, on a limited sample set to get a first look on what information there is in EAS CL distributions, so as to have a starting point for comparison during detector optimization.

### 5.1. Energy Estimation

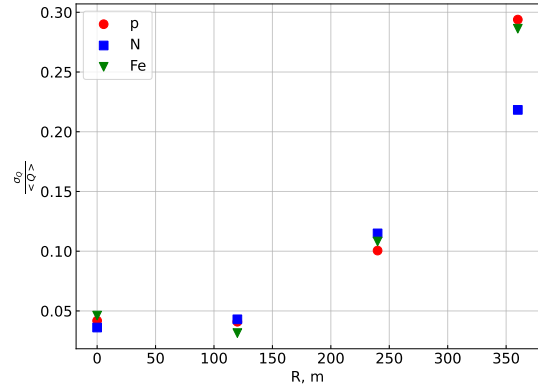
Energy estimation is possible using only reflected CL data. Direct CL registered with a small scale detector without information on axis position relatively to the detector does not allow for such a procedure.

At first, simple approach, the total number of photons  $Q$  reaching the detector SiPM mosaic was used as the main criteria for energy estimation. This number depends on the primary particle energy  $E_0$  and the distance from the center of the detector field of view to the shower axis  $R$ . Such dependencies (i.e.  $Q(E_0; R)$ ) can be obtained as a regression over the precalculated model values for different parameters (energies, angles, atmosphere conditions). Energy estimations are then based on  $Q^{exp}$  and  $R^{exp}$  as:  $E_0^{est} = E_0(Q^{exp}, R^{exp})$ .

An example of a modelled  $Q(E_0; R)$  dependence is shown in Fig. 10 for 1000 m detector flight altitude. Five model EAS with a 10 PeV primary energy for different nuclei with 100 random axis locations per shower (e.g. 500 data points per nuclei) were used. The relative fluctuations  $\sigma Q$  (see Fig. 11), as expected, grow with distance as the total amount of light collected on the detector gets smaller. The resulting uncertainty inevitably will play major role in the overall energy estimation error.



**Figure 10.** Total number of photons collected from a 10 PeV shower depending on the distance from the shower axis to the center of the detector field of view. Red dots represent data from primary protons, blue – nitrogen nuclei, green – iron nuclei

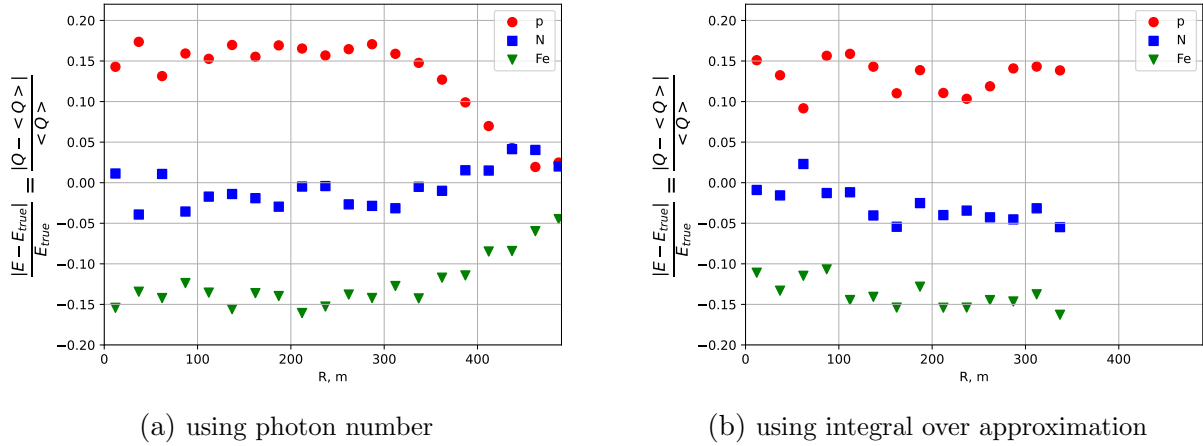


**Figure 11.** The relative fluctuations of the total number of reflected EAS CL photons at several distances for the showers from the left panel. Red dots represent data from primary protons, blue – nitrogen nuclei, green – iron nuclei

Since the primary particle mass is unknown in the experiment prior to a specific analysis, it is common practice to use average EAS characteristics with corrections applied at a later stage. If such an average (shown as a magenta line in Fig. 10) is applied to the same sample set of 10 PeV showers the resulting errors will not exceed 17% (see Fig. 12a).

This simple approach shows good ability to estimated primary mass, however, it can be updated for higher precision. The main issue with simple approach is that after certain distance  $R$  the significant portion of the shower spot is outside the detector field of view. To solve this issue an approximation of the photons LDF can be used to reconstruct their density. Thus, the photons distribution was fitted with a rational function (same as in [17]):

$$I(r) = \frac{p_0^2}{(1 + p_1 r + p_2 r^2 + p_3 r^{1.5})^2 (1 + p_4 r^s)}, \quad (2)$$



**Figure 12.** Errors in energy estimations for different nuclei based on the average  $E(Q; R)$  dependence for two approaches (see text). On the left,  $E(Q; R)$  was obtained using total number of photons in the detector. On the right, same procedure was done using LDF approximation. Red dots represent data from primary protons, blue – nitrogen nuclei, green – iron nuclei. Nitrogen has lower estimation errors since it has the closest to average profile

where  $r$  is the distance from shower axis,  $p_i$  and  $s$  are free parameters. An integral over this smooth function gave a better estimation of total number of photons in the shower and a bit better results. The errors in the primary energy reconstruction became smaller and did not exceed 15% (see Fig. 12b). It should be noted that approximation procedure has its cost and at the moment is not perfectly stable, up to 30% of the LDF fits failed. However, this happened mostly at lower energies (5 PeV showers were also considered) and at high  $R$ , where the number of available data points for approximation is low, what was somewhat expected.

The algorithm will undergo further improvements, namely, we have to optimize the set of limitations imposed on the registered events in order to make procedure more reliable (and reduce the number of lost events), while keeping the energy estimate error low.

## 5.2. Mass Estimation

Contrary to the energy, EAS primary particle mass can be estimated based on both direct and reflected CL.

### 5.2.1. Mass estimation using reflected Cherenkov light

Primary particle mass estimation was done using the same approach as previously used for the SPHERE-2 experiment [17]. In general, the CL lateral distribution function  $I(r)$  correlates with shower longitudinal profile, therefore, there may exist a parameter in the mentioned lateral distribution form description that should be sensitive to the primary particle mass. And such a parameter was found – integral steepness  $\eta$ , i.e. defined by the major part of the distribution and not just a single value at some point:

$$\eta = \frac{\int_0^{r_1} I dr}{\int_{r_1}^{r_2} I dr}, \quad (3)$$

where  $r_i$  are distances from shower axis. These distances can be selected according to the desired property of the shower separation criterion. In our case, the target property was minimal errors in shower separation by primary particle mass.

The steepness  $\eta$  is defined through the CL lateral distribution function  $I$  which traditionally is defined over the observation level. In our studies, we chose to use the CL distribution over the SiPM mosaic, a projection of the original distribution, known at the SiPM positions. Also, the real data will contain background and statistical fluctuations. Thus, the distribution was fitted with a rational function (same as in section above).

So, the mass estimation procedure now consists of shower data fitting and calculations of parameter  $\eta$ . This procedure was done for a set of model EAS from 10 PeV primary protons, nitrogen and iron nuclei with small zenith angles. For each EAS the parameter  $\eta$  was estimated for some set of  $r_i$  parameters. The shower separation procedure over  $\eta$  was done using some varying thresholds. Quality was then evaluated and the  $r_i$  set was updated. In the end the optimal set of  $r_i$  and thresholds was obtained with the following shower separation quality:

**Table 1.** Shower separation quality for p-N and N-Fe pairs

class	p-N	N-Fe
border value	0.699	0.614
class error	0.314	0.317

These results were obtained only for one energy and without background so far, but further investigations are planned for future work.

### 5.2.2. Direct Cherenkov light

The primary particle mass estimation from direct CL data was studied in two separate ways – from clean CL distributions themselves and from separate detector data. Mass estimation using direct CL data from the detector SiPM mosaic is not yet finished.

Direct CL distributions used the same data set as for the arrival direction (see section 4.2.2). Since the detector “sees” EAS from the side the direct CL spot is elongated and rotated in the direction of the shower axis. For our analysis, this angle can be found as:

$$\tan(2\varphi) = \frac{2\sigma_{xy}}{\sigma_{xx} - \sigma_{yy}}, \quad (4)$$

where  $\sigma_{xy}, \sigma_{xx}, \sigma_{yy}$  are second central momenta of the CL distribution ( $x$  and  $y$  are orthogonal coordinates fixed to the detector). The spot’s major and minor axes lengths then can be estimated as:

$$a_1 = \sigma_{xx} \cdot \cos^2(\varphi) + 2\sigma_{xy} \cdot \sin(\varphi) \cdot \cos(\varphi) + \sigma_{yy} \cdot \sin^2(\varphi), \quad (5)$$

$$a_2 = \sigma_{yy} \cdot \cos^2(\varphi) - 2\sigma_{xy} \cdot \sin(\varphi) \cdot \cos(\varphi) + \sigma_{xx} \cdot \sin^2(\varphi), \quad (6)$$

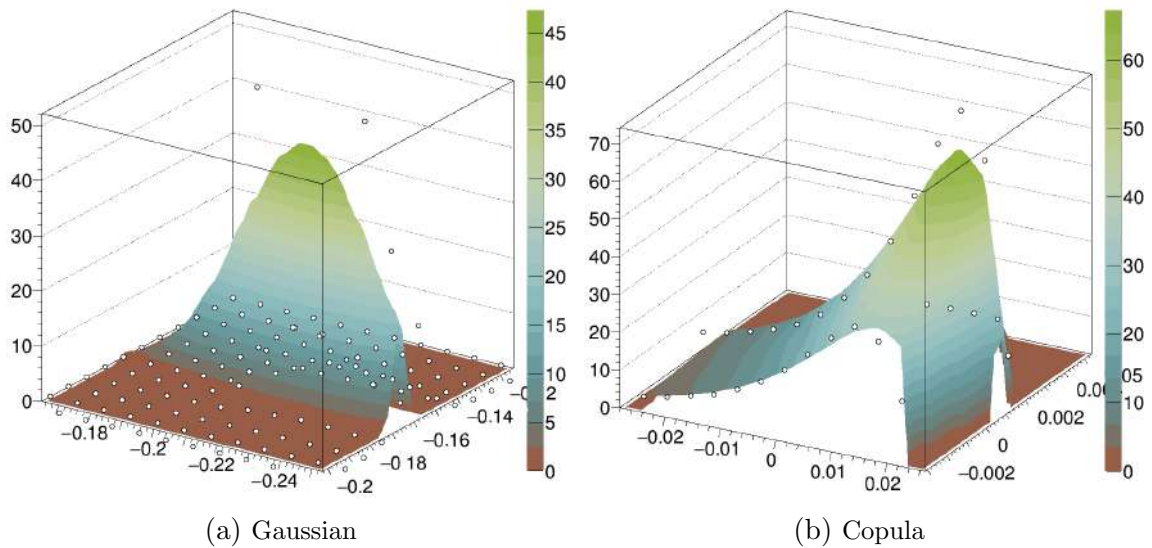
The major axis length is equal to  $a = \max\{a_1, a_2\}$  and can be used as the primary particle mass separation criterion. The precision of  $\varphi$  and  $a$  estimation by this approach is sensitive to the noise level so several thresholds were tested (e.g. rejection of everything below a certain value) both absolute and relative to the maximum of the distribution.

Another option was to fit the spot with some shape function (see Fig. 13a for an example):

$$F(x', y') = p_0 \cdot \exp \left[ -\frac{(x' - p_1)^2}{p_2^2} - \frac{(y' - p_3)^2}{p_4^2} - \frac{2p_5 \cdot (x' - p_1) \cdot (y' - p_3)}{p_2 \cdot p_4} \right], \quad (7)$$

where  $x'$  and  $y'$  are the rotated by  $\varphi$  coordinates,  $p_i$  are free parameters. The spot's long axis can be estimated as  $a = \max\{p_2, p_4\}$ .

Comparison of the quality of separation using estimations of the direct CL spot's long axis by these two approaches are show in Tab. 2. The results of approximation are not very good, since the distribution (7) is symmetrical along its major axes and the direct CL spot is definitely not. So, a new function is now in test – normal distribution over the short axis coupled with a Gamma-distribution along the long axis (see Fig. 13b for an example).



**Figure 13.** Sample fit functions for direct CL lateral distribution function: asymmetrical normal distribution on the left and normal copula function on the right

**Table 2.** Primary particle separation errors ( $\sigma$ ) using direct CL distribution parameters and lens detector data, simulations used bin size  $0.5^\circ \times 0.5^\circ$ , presented values have typical uncertainty around 0.02

approach	CL distribution				lens detector			
	p-N		N-Fe		p-N		N-Fe	
	$\sigma_p$	$\sigma_N$	$\sigma_N$	$\sigma_{Fe}$	$\sigma_p$	$\sigma_N$	$\sigma_N$	$\sigma_{Fe}$
abs. 3 photons	0.25	0.24	0.24	0.26	0.32	0.32	0.32	0.32
abs. 5 photons	0.27	0.26	0.24	0.26	0.27	0.27	0.27	0.27
rel. 5% max	0.35	0.29	0.29	0.29	0.49	0.49	0.47	0.47
rel. 7% max	0.35	0.19	0.29	0.29	0.36	0.35	0.38	0.38
approximation	0.62	0.12	0.23	0.27	–	–	–	–

This analysis was also done for a simple direct CL detector (one lens, large sensor) on the same data sample. While for the arrival direction estimations the image distortions are present,



but can be easily corrected, distortion of the spot shape does affect the mass separation quality. The analysis above applied to the simple detector data showed significantly larger errors (see Tab. 2).

## Conclusion

Here we described an ongoing development process of the SPHERE-3 telescope aimed at energy spectrum and mass composition study of primary cosmic rays in the 1–1000 PeV energy range. In comparison with the previous SPHERE detectors, which registered reflected CL from extensive air showers, the new one will also register direct CL from the same shower providing more valuable data on primary particle parameters. The detector design is still in progress, methods for the detection of direct CL are being tested along with the data analysis approaches.

The updated algorithms for processing reflected light data and the parallelized variant of the CORSIKA code which will allow to model the necessary characteristics of CL are considered.

The considered design of the reflected light telescope makes it possible to estimate the direction and mass of the primary particle no worse than the SPHERE-2 telescope. Its maximum error in estimating the primary energy is evaluated.

It is found that the shower arrival direction of the shower can be defined by the angular distribution of direct light with an error not exceeding  $0.5^\circ$ .

It is proved that the angular distribution of CL is sensitive to the mass of the primary particle and using the length of the spot's long axis as a criterion the primary particles can be divided by mass with classification errors equal or lower than when using CL reflected from the snow.

Also, it was realized that for a better study of the image one should approximate it by a function different from the two-dimensional Gaussian distribution and is asymmetric.

These first results were obtained using a simulations pipeline described here that will be used as the foundation of the general experiment optimization loop. Such optimization is needed due to the unique approach of the SPHERE experiments and, therefore, natural lack of readily available methods and tested solutions. The optimization, however, will be extremely computationally heavy with more and more parameters being set loose in the process. Right now, the only free optimization parameters available are the thresholds for the procedures, however, in the future, it is planned to set detector geometry and other parameters (such as flight altitudes and detector operation regimes) as free parameters in the optimization process, given it will be computationally possible.

## Acknowledgements

The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [21].

This work is supported by the Russian Science Foundation under Grant No. 23-72-00006, <https://rscf.ru/en/project/23-72-00006/>.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Agostinelli, S., Allison, J., Amako, K., *et al.*: Geant4—a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506(3), 250–303 (2003). [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8)
2. Allison, J., Amako, K., Apostolakis, J., *et al.*: Geant4 developments and applications. IEEE Transactions on Nuclear Science 53(1), 270–278 (2006). <https://doi.org/10.1109/TNS.2006.869826>
3. Allison, J., Amako, K., Apostolakis, J., *et al.*: Recent developments in Geant4. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 835, 186–225 (2016). <https://doi.org/10.1016/j.nima.2016.06.125>
4. Antonov, R.A., Aulova, T.V., Bonvech, E.A., *et al.*: Detection of reflected Cherenkov light from extensive air showers in the SPHERE experiment as a method of studying super-high energy cosmic rays. Phys. Part. Nucl. 46, 60–93 (2015). <https://doi.org/10.1134/S1063779615010025>
5. Antonov, R.A., Beschapov, S.P., Bonvech, E.A., *et al.*: Results on the primary CR spectrum and composition reconstructed with the SPHERE-2 detector. Journal of Physics: Conference Series 409 (feb 2013). <https://doi.org/10.1088/1742-6596/409/1/012088>
6. Bakhromzod, R., Galkin, V.: The search and analysis of optimal criteria for the selection of extensive air showers from  $\gamma$ -quanta by Cherenkov telescopes. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 1018 (2021). <https://doi.org/10.1016/j.nima.2021.165842>
7. Bonvech, E.A., Azra, C., Chernov, D.V., *et al.*: Design of the Simulation Scheme for SPHERE-3 Telescope for the  $10^{15}$ – $10^{18}$  eV Primary Cosmic Ray Studies Using Direct and Reflected Cherenkov Light from the Extensive Air Showers. Phys. Atom. Nucl. 86(6), 1048–1055 (2023). <https://doi.org/10.1134/S1063778824010149>
8. Budnev, N., Chernov, D., Gress, O., *et al.*: Tunka-25 Air Shower Cherenkov array: The main results. Astropart. Phys. 50-52, 18–25 (2013). <https://doi.org/10.1016/j.astropartphys.2013.09.006>
9. Chernov, D.V., Azra, C., Bonvech, E.A., *et al.*: SPHERE-3 Project for Studying the Composition of Primary Cosmic Rays in the Energy Range Between 1 and 1000 PeV. Phys. Atom. Nucl. 85(6), 641–652 (2022). <https://doi.org/10.1134/S1063778822060059>
10. Chudakov, A.: A possible method to detect EAS by the Cherenkov radiation reflected from the snowy ground surface. (in russian). In: Experimental methods of studying cosmic rays of superhigh energies: Proc. All-Union Symposium. vol. 620, pp. 69–72 (1972)
11. Galkin, V.I., Borisov, A.S., Bakhromzod, R., *et al.*: A method for estimation of the parameters of the primary particle of an extensive air shower by a high-altitude detector. Moscow University Physics Bulletin 73(2), 179–186 (2018). <https://doi.org/10.3103/S0027134918020078>

12. Galkin, V.I., Gzhatdov, T.A.: Classifying groups of PCR nuclei with energies of  $10^{15}$ - $10^{16}$  eV according to the spatial-angular distribution of EAS Cherenkov light. Bulletin of the Russian Academy of Sciences: Physics 75(3), 309–312 (Mar 2011). <https://doi.org/10.3103/S1062873811030166>
13. Galkin, V., Borisov, A., Bakhromzod, R., *et al.*: EAS primary particle parameter estimation with the complex Pamir-XXI detector array. EPJ Web Conf. 145 (2017). <https://doi.org/10.1051/epjconf/201614515004>
14. Heck, D., Knapp, J., Capdevielle, J.N., *et al.*: CORSIKA: A Monte Carlo code to simulate extensive air showers. Report FZKA-6019 (2 1998). <https://doi.org/10.5445/IR/270043064>
15. Kalmykov, N., Ostapchenko, S., Pavlov, A.: Quark-gluon-string model and EAS simulation problems at ultra-high energies. Nuclear Physics B - Proceedings Supplements 52(3), 17–28 (1997). [https://doi.org/10.1016/S0920-5632\(96\)00846-8](https://doi.org/10.1016/S0920-5632(96)00846-8)
16. Knurenko, S., Petrov, I.: Mass composition of cosmic rays above 0.1 EeV by the Yakutsk array data. Advances in Space Research 64(12), 2570–2577 (2019). <https://doi.org/10.1016/j.asr.2019.07.019>
17. Latypova, V.S., Nemchenko, V.A., Azra, C.G., *et al.*: Method for Separating Extensive Air Showers by Primary Mass Using Machine Learning for a SPHERE-Type Cherenkov Telescope. Moscow University Physics Bulletin 78(1), S25–S31 (Dec 2023). <https://doi.org/10.3103/S0027134923070196>
18. Ostapchenko, S.: LHC data on inelastic diffraction and uncertainties in the predictions for longitudinal extensive air shower development. Phys. Rev. D 89 (Apr 2014). <https://doi.org/10.1103/PhysRevD.89.074009>
19. Podgrudkov, D.A., Bonvech, E.A., Vaiman, I.V., *et al.*: First results from operating a prototype wide-angle telescope for the TAIGA installation. Bulletin of the Russian Academy of Sciences: Physics 85(4), 408–411 (Apr 2021). <https://doi.org/10.3103/S1062873821040286>
20. Poole, C.M., Cornelius, I., Trapp, J.V., Langton, C.M.: A CAD interface for GEANT4. Australasian Physical & Engineering Sciences in Medicine 35(3), 329–334 (Sep 2012). <https://doi.org/10.1007/s13246-012-0159-8>
21. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. Supercomputing Frontiers and Innovations 6(2), 4–11 (Jun 2019). <https://doi.org/10.14529/jsfi190201>