

Ultra-high energy air shower simulation without thinning in CORSIKA

Tanguy Pierog^{*a}, R. Engel^a, D. Heck^a, G. Poghosyan^b, J. Oehlschläger^a, and D. Veberič^a

^aKarlsruhe Institute of Technology (KIT), IKP, D-76021 Karlsruhe, Germany

^bSimulation Lab E&A Particles, SCC, D-76021 Karlsruhe, Germany

E-mail: tanguy.pierog@kit.edu

Interpretation of EAS measurements strongly depends on detailed air shower simulations. One of the severe limitations is the calculation time of Monte-Carlo programs like CORSIKA at very high energies. Thinning algorithm has been introduced in the past to reduce the computation time and disk space of the output at the price of the loss of small scale structures in simulated air showers. Thanks to the newly developed parallelization scheme and special tools to study multiple thinning levels for a given shower on a limited disk space, it is now possible to compare thinned and unthinned simulations of a single shower to quantify these losses. Preliminary results will be presented together with the details of the latest release of CORSIKA.

*The 34th International Cosmic Ray Conference
30 July - 6 August, 2015
The Hague, The Netherlands*

*Speaker.

1. Introduction

The experimental method of studying ultra-high energy cosmic rays is an indirect one. Typically, one investigates various characteristics of extensive air showers (EAS), a huge nuclear-electromagnetic cascade induced by a primary particle in the atmosphere, and uses the obtained information to infer the properties of the original particle, its energy, type, direction etc. Hence, the reliability of ultra-high energy cosmic ray analyzes depends on the use of proper theoretical and phenomenological descriptions of the cascade processes.

The most natural way to predict atmospheric particle cascading in detail seems to be a direct Monte Carlo (MC) simulation of EAS development, like it is performed, for example, in the CORSIKA program [1]. As very large computation times are required at ultra-high energy, a parallelization of the sub-showers of the cascade on large CPU clusters is now possible but creates other problems with the size of the output file.

In this article, we will present a new option to reduce the size of the data file produced by CORSIKA and a new way to study the effect of the use of weighted particles (thinning) with a direct comparison of the corresponding fully simulated shower. First results will be shown at 10^{19} eV and 10^{20} eV.

2. Parallel calculation

Since CORSIKA 7.3x a possibility to run unthinned showers using massive parallelization with Message Passing Interface (MPI) management has been introduced.

To overcome the extreme long computing time for a single shower induced by an ultra-high energy primary particle the simulation task has to be split into many jobs which are treated in parallel on many cores of a computer cluster. In our approach the results of such parallel simulations are controlled in a unique way by seeds for the random number generator given by the user. It is now possible to collect all particles above a user-defined threshold in an external file, which can be used to run all the sub-showers induced by these particles on different CPUs. If the seeds are well defined for each sub-shower, it is possible to reproduce the same shower under different technical conditions of the available computing resources. A challenge is the merging and handling of the huge final file containing all particles which arrive at the ground.

To help book-keeping and for a better management of all the sub-showers on large clusters, there is the possibility to use CORSIKA as a subroutine of a master program which can distribute all the sub-showers on different CPUs using MPI protocol. In that case, there is no need for an external file to save temporary stack. To reduce the very large number of files produced by all the subshowers, output data are collected and saved in a single file per CPU core.

As a side effect, an option has been added to introduce a particle or a list of particles anywhere in space independently of the shower axis. This can be used to study artificial showers having special features.

3. Unthinned shower

Thanks to the parallelization of the calculation, it is now possible to get in few hours the equivalent of 6 months or one year of single-core computing time. Hence EASs with energy as high

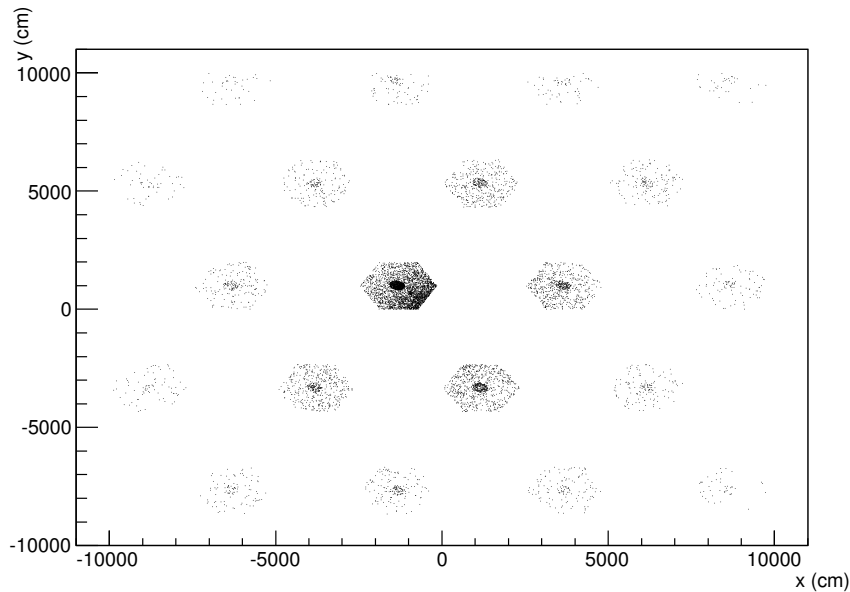


Figure 1: Particles at ground with AUGERHIT option for a proton induced shower with an energy of 10^{16} eV and a zenith angle of 38° . The band size is 10 m with a spacing of 50 m.

as 10^{20} eV can be generated without the use of any method to artificially reduce the simulation time, like the so-called thinning. This is what we call “unthinned shower” in which, by definition, each particle has a weight of 1. As a consequence, if all particles reaching the ground are recorded, the list of these particles is larger than a Terabyte (TB). Such amount of data is of course challenging to manage and some solutions have to be found to solve this problem. Nevertheless, one of the first things that can be done with an unthinned shower is to compare it to what was used until now: Thinned showers where weights are given to particles to compensate for the fact that some particles were discarded during the simulation to save time and disk space.

3.1 AUGERHIT option

To reduce the size of the CORSIKA output file (DAT file) the usage of this file has to be taken into account.

Since experimentally ultra-high energy EAS can only be studied at the ground using sparse array detectors like the Pierre Auger Observatory [2], in CORSIKA we introduced a new option AUGERHIT [3] which writes to disk only particles falling into a hexagonal grid whose parameters can be defined by the user. An example of such a grid is shown in Fig. 1 with a band size of 10 m and a spacing of 50 m.

In case of an unthinned shower, even such sparse geometry can lead to relatively large DAT files. To reduce even further the number of particles saved to disk, an additional option allows to save only particles which hit a water tank as used for the Pierre Auger Observatory, taking into account the full 3D geometry. The so called “tank shadow” can be seen in Fig. 1 as the darker points in the middle of the hexagons. To use the same shower for different core positions within the grid, it is possible to have up to 20 different positions in one output file.

Since real detectors have a limited dynamic range, they will always saturate close to the shower core because of the huge density of particles. So it is optionally possible to suppress the particle recording within a circle around the core with a user defined radius.

Using these different options, it is possible to reduce the output DAT file from many TB to around 2 Gigabyte (GB) for a shower with primary energy of 10^{20} eV.

3.2 MULTITHIN option

Since air shower simulations reach a relatively high energy (above 10^{17} eV), computation time has been considered as a limiting factor to build up air shower libraries with large enough statistics necessary for cosmic ray experiment analysis. As a consequence the Hillas thinning algorithm [4] has been introduced very early on. The principle is very simple: When the number of secondary particles is high enough (energy typically around the fraction $T = 10^{-6}$ of the primary energy) one can conserve only one randomly selected particle after each interaction (hadronic or electromagnetic) of a particle with energy E_i . To conserve the total energy in the shower one can give to this particle with energy E_p a weight W_p such that we have $E_i = W_p E_p$. By construction, on an event-by-event basis the energy is always conserved, but not the number of particles. For a large number of simulations the average values are well reproduced [5] but some artificial fluctuations are introduced. When the value of the threshold T is too large, these fluctuations can become very important. To limit this effect, a maximum weight can be introduced [6].

The effect of thinning has been studied in [5] at energies where it was still possible to run large statistics of unthinned showers. But at energies as high as 10^{19} or 10^{20} eV, in particular because even if a shower starts with the same first interaction, it is impossible to have exactly the same shower with and without thinning (different number of particles therefore different random number chain).

With the development of parallel calculation allowing the simulation of unthinned showers at ultra-high energy, a new option has been introduced in CORSIKA. MULTITHIN [7] makes it possible to get in a single run all the particles without weight and in addition (in a new data block) 6 weights corresponding to 6 different thinning levels (including different random numbers for the same parameters) which are computed as usual but instead of discarding the rejected particle, they get a negative weight as a flag not to be used in another weight calculation. As a consequence, the large majority of particles reaching the ground (with weight 1) have the additional weight list full of negative numbers (particles discarded by thinning) and a few of them have some positive value which can then be used to reconstruct a thinned event. So event-by-event the very same simulation can be analyzed with or without thinning.

Since this option doubles the size of the DAT file (one additional weight block for each particle block), it can be combined with the AUGERHIT option to reduce the output file. In this case all particles hitting the tank shadow are recorded but only the particles with at least one positive weight (corresponding to some thinned shower) are recorded in the hexagonal grid.

3.3 Results

In this section we will present energy (Fig. 2) and time (Fig. 3) distributions of γ at 500 m, e^- at 1000 m and μ^- at 2000 m from the shower core from a proton induced shower with a zenith

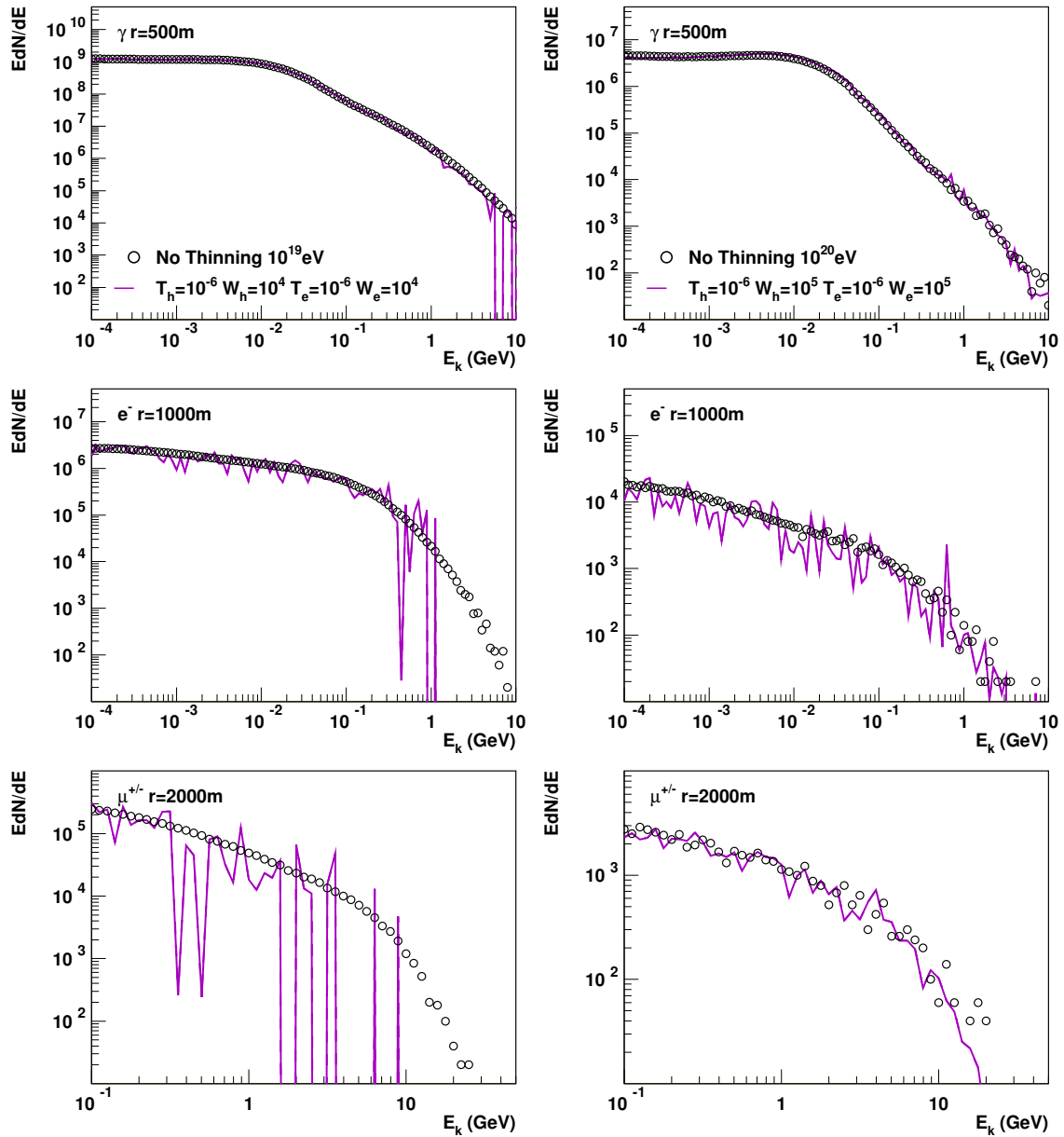


Figure 2: Energy distributions for γ at 500 m, e^- at 1000 m and μ^- at 2000 m from the core. See text for details.

angle of 38° at the Pierre Auger Observatory site (1492 m a.s.l.). On the left-hand side of the figures the primary energy is 10^{19} eV and all particles reaching the ground at a distance larger than 200 m from the shower core are recorded. The DAT file has a size of 670 GB. It will be defined as shower 1. On the right-hand side the primary energy is 10^{20} eV and particles reaching the ground at a distance larger than 200 m from the shower core are recorded using the AUGERHIT option with a grid size of 750 m and a radius of 35 m. 20 different core positions are recorded and the DAT file has a size of 1.7 GB. It will be defined as shower 2.

For all plots the shower without thinning is presented by black circles for the energy distri-

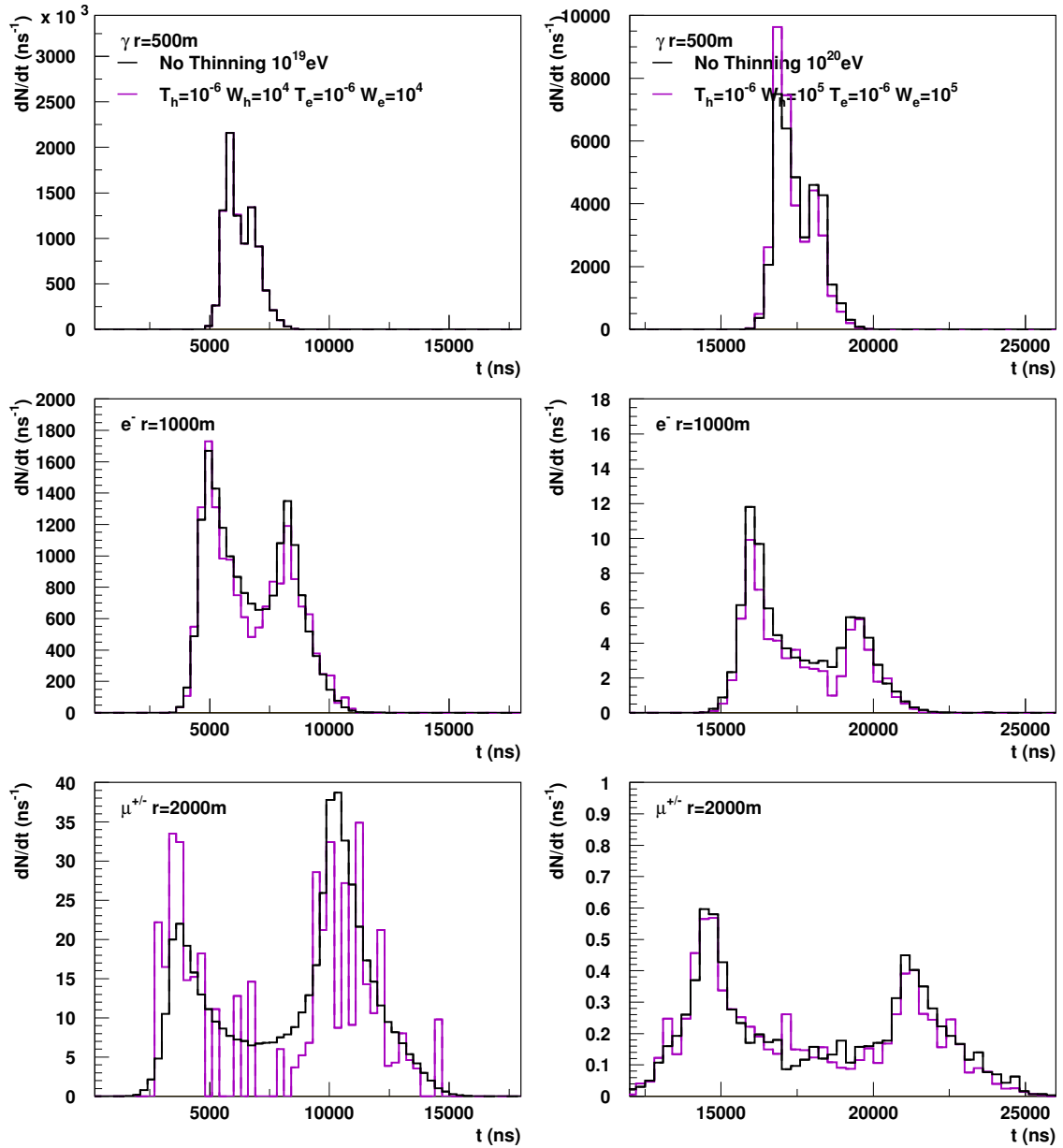


Figure 3: Time distributions for γ at 500 m, e^- at 1000 m and μ^- at 2000 m from the core. See text for details.

butions and black histograms for the time distribution. One selected thinning level is shown as a purple line. For both electromagnetic and hadronic particles, the thinning threshold is $T = 10^{-6}$ and both hadronic and electromagnetic maximum weights are fixed to $W_{\max} = T E_{\text{primary}}/\text{GeV}$ (10^4 at 10^{10} GeV and 10^5 at 10^{11} GeV). The large difference in the number of particles between the two different showers is due to the fact that particles are counted in a ring around 500 m, 1000 m or 2000 m without normalization by the surface. As a consequence for shower 1, particles are recorded for the complete rings while for shower 2 only the tank shadow included in these rings is taken into account (and for the thinned shower, the surface of the hexagon has been renormalized to

the surface of the tank shadow to compare with the unthinned shower). The two peaks observed in the time distribution are due to the fact that the observation plane is not perpendicular to the shower axis. As a consequence, the particles upstream arrive earlier compared to the particle downstream. It is a simple geometric effect.

From the energy distributions we can conclude as expected that when the number of particles is large (like for γ 's at 500 m) the result of the thinned shower is completely equivalent to the unthinned shower. When the maximum weight is relatively large compared to the total number of particles like electrons, the artificial fluctuations increase, but the average is still well reproduced. But when the number of particles becomes lower than the maximum weight (like for muons at 2000 m in shower 1, Fig. 2 lower-left corner), then fluctuations are very large and we barely recognize the original distribution which is still smooth for the unthinned shower.

For the time distributions, the situation is slightly different. When the number of particles is extremely high (as in the upper-left corner of Fig. 3), the thinned and unthinned showers give the same result. But otherwise, time distributions are changed by the thinning process and in particular far from the core some artificial peaks can appear. Therefore any analysis based on time traces may be biased by thinning. Further studies with higher time resolution and with full detector simulation will be performed to test the effects of thinning and resampling [8] (used to generate particles without weight out of particles with weight) on various analyzes used at the Pierre Auger Observatory.

4. Summary and outlook

The next version of the air shower simulation code CORSIKA will be released by the end of the year 2015. One of the main changes is the improvement of parallel calculation using MPI for air shower simulations on thousands of CPUs. Two new options will be available to reduce the output data file produced by the ultra-high energy unthinned showers and to study in details the effects of thinning algorithms on an event-by-event basis. First results show that thinning does not introduce bias when a large number of particles are considered at the ground. This condition is not being satisfied for small detectors at large distances from the core. Further studies have to be performed to check the effects of resampling based on weighted particles in particular for time distributions.

Other new features in CORSIKA include a new version of the hadronic interaction model SIBYLL which now includes charm production [9], which is particularly important for high energy neutrino production. With the new ICECUBE option it is now possible to stop a shower simulation as soon as there is no chance to produce high energy neutrinos in this particular event. It reduces the production time of libraries, keeping only events useful for neutrino experiments.

For parallel calculation, the next step will be to optimize the distribution of CPU cores as a function of the shower energy to be able to submit a large number of shower simulations with different energies in a single job. Indeed supercomputers with thousands of cores allow only parallel calculation from a unique job like for one unthinned ultra-high energy air showers. As a consequence it is necessary to develop a new algorithm to optimally distribute the generation of multiple showers of various energy to be able to use such supercomputers offering large time of computing power. Instead of generating one 10^{20} eV shower in 10 hours, 5000 CPUs could be use simultane-

ously to generate 1000 showers of 10^{17} eV in 10 hours for instance.

Acknowledgements

The CORSIKA authors would like to thank J. van Santen and K. Jero (IceCube collaboration) for the development of the ICECUBE option, and all users who helped to detect and correct errors in the code.

References

- [1] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, *Forschungszentrum Karlsruhe, Report FZKA 6019* (1998).
- [2] **Pierre Auger** Collaboration, A. Aab et al., *The Pierre Auger Cosmic Ray Observatory*, *accepted for publication in Nucl. Instrum. Meth. A* (2015) [[arXiv:1502.0132](https://arxiv.org/abs/1502.0132)].
- [3] J. Oehlschläger and D. Heck, *The AUGERHIT option of the air shower simulation program CORSIKA*, *KIT Scientific Working Papers KIT-SWP 20* (2014).
- [4] A. M. Hillas, *Two Interesting Techniques for Monte Carlo Simulation of Very High Energy Hadron Cascades*, *Proc. 17th Int. Cosmic Ray Conf., Paris 8* (1981) 193.
- [5] D. Gorbunov, G. Rubtsov, and S. V. Troitsky, *Air-shower simulations with and without thinning: Artificial fluctuations and their suppression*, *Phys. Rev. D 76* (2007) 043004, [[astro-ph/0703546](https://arxiv.org/abs/astro-ph/0703546)].
- [6] M. Kobal, *A thinning method using weight limitation for air-shower simulations*, *Astropart. Phys. 15* (2001) 259–273.
- [7] D. Heck, *The MULTITHIN option of the air shower simulation program CORSIKA*, *KIT Scientific Working Papers KIT-SWP 17* (2014).
- [8] P. Billoir, *A sampling procedure to regenerate particles in a ground detector from a 'thinned' air shower simulation output*, *Astropart. Phys. 30* (2008) 270–285.
- [9] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *A new version of the event generator SIBYLL*, *34th Int. Cosmic Ray Conf., The Hague, The Netherlands, 30 July - 6 August, 2015*.