

Study of statistical thinning with fully-simulated air showers at ultra-high energies

R. Bruijn*, F. Schmidt†, J. Ilce‡, and J. Knapp*

* School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, United Kingdom

† Department of Astronomy and Astrophysics, University of Chicago

‡ School of Physics and Astronomy, University of St. Andrews

Abstract. The full simulation of extensive air showers at high energies requires enormous computing times. Therefore, statistical thinning techniques are usually applied, which speed up computing by $> 10^4$. Adverse effects of thinning on air shower observables are enhanced fluctuations and systematic shifts. The CORSIKA program has been extended to produce fully simulated showers at ultra-high energies, by parallelisation of the calculation, and to study of the effect of thinning, by simultaneous simulation of the “unthinned” and “thinned” shower. Here results on thinning and fluctuations in a vertical 10^{19} eV proton shower are presented and the uncertainties in typical air shower observables are discussed.

Keywords: extensive air shower, simulation, statistical thinning

I. INTRODUCTION

The simulation of extensive air-showers plays a key role in the interpretation of data measured with air shower arrays. Simulations become difficult for experiments which aim to study showers at ultra-high energies ($> 10^{19}$ eV). At such high energies, showers contain in excess of 10^{10} secondary particles and moreover, multiple showers must be simulated to account for the intrinsic shower-to-shower fluctuations. The full simulation of such showers, which involves tracking all individual particles and interactions, would require a prohibitively large amount of computer time and storage space. To overcome this limitation, *statistical thinning* [1], [2] was introduced. Only a representative subset of particles is tracked, the others being discarded. The tracked particles are assigned an appropriate statistical weight (w) to account for the energy of the discarded particles. Thinning reduces the processing time to a manageable level and preserves, on average, the observables in the shower, but it also introduces artificial fluctuations which cause biases in parts of the shower where only small particle numbers are present. Also, correlations between particles can get lost, e.g. as electromagnetic particles are thinned separately from muons, a correlation between the muon and its electromagnetic halo can disappear. Therefore, a balance is needed between gain in computing time and the level of detail maintained in the simulated shower.

Two parameters used in the thinning algorithms are the thinning level (ε), which specifies at what fraction of the primary energy thinning sets in, and a maximum weight (w_{\max}), which prevents the particle weights from becoming too large and thus producing excessive artificial fluctuations. Due to the large difference in numbers between electromagnetic particles (photons, electrons and positrons) and hadrons and muons, the thinning levels and maximum weights are set separately for each species.

In every experiment, simulations of a detailed response of a detector to the shower particles are needed. The statistical weights are here problematic. This requires first the smoothing of the thinned particle distribution to reproduce a set of unweighted particle falling on the detector. A procedure to do this *unthinning* is given in [3]. In this procedure, all (thinned) particles arriving in an area A_{samp} , larger than the area of the detector element A_{det} , are used to represent the particles accounted for by the weights and to calculate the detector response. Ideally, the $A_{\text{samp}}/A_{\text{det}}$ should be approximately equal to the average particle weight.

Typically for these high-energy showers, measurements at large distances from the shower core (> 500 m) are used to estimate shower properties such as energy, arrival direction and depth of shower maximum. At large distances, however, the particle densities become increasingly sparse and thus potentially sensitive to the detrimental effects of statistical thinning.

II. MODIFICATIONS TO CORSIKA

To study the effects of thinning, it is necessary to simulate a single shower unthinned and, simultaneously, with different thinning levels. This separates the fluctuations introduced by the thinning from the physical shower fluctuations when simulating different showers, as the random number sequences differ as soon as a particle is discarded at a given thinning level. For this study, version 6.720 of CORSIKA [4] has been extended in two ways. CORSIKA has been extended with an option (“multi-weights”) to simulate a shower simultaneously, unthinned and thinned, with up to 10 different thinning levels. Consequently, each particle carries multiple weights, one weight for each of the thinning methods. For each particle these weights indicate whether the particle would make it to the ground ($w > 0$) or whether

Primary particle	Proton
Energy	10^{19} eV
Zenith angle	0°
High-energy hadr. interactions	QGSJET II
Low-energy hadr. interactions	Fluka 2006.3b
Depth (height) of first interaction	64.47 g/cm^2 (19.197 km)
Observation level	870 g/cm^2 (1452 m)
Radial thinning	< 200 m
Energy cuts (photons, electrons)	0.1 MeV
Energy cuts (muons, hadrons)	0.1 GeV

TABLE I
Main input parameters for the unthinned shower.

it would have been cut ($w = -1$) by a given thinning method. All (thus unthinned) particles which reach the observation level are stored with all their weights. When analysing a particular thinning level, only the particles with that specific weight > 0 are considered.

The second extension provides a facility to parallelise the calculation of a shower for time reduction by distributing the simulation over multiple CPUs. This is implemented by writing out particles (and discarding them) if the energy is below a set high threshold. These particles serve subsequently as starting particles for subshowers that can be distributed on many processors.

III. SIMULATION OF A SHOWER

This extended version of CORSIKA has been used to study (after simulation) an unthinned proton-induced shower of 10^{19} eV. The shower simulation parameters are summarized in table I. The choice of a vertical shower ensures an azimuthal symmetry around the shower axis to allow a study of fluctuations. The depth of first interaction was set to about 64 g/cm^2 , the average point of interaction for an 10^{19} eV proton. The starting random number was chosen to produce a first interaction with an average inelasticity. The depth of shower maximum is 799 g/cm^2 . Tab. II gives the details of the thinning methods used. Options 1 to 5 correspond to *optimum thinning* according to [2] (i.e. with a weight limitation that minimizes the artificial fluctuations for the CPU time used) and are the main focus of this study.

In an initial run the shower was simulated until the energy of the secondary particles fell below 10^{16} eV. This yielded a total of 1241 sub-showers which then were processed individually on a multiprocessor farm. While the total CPU time used was about 1.5 years, it took only 7.5 days to complete the simulation by running the sub-showers in parallel, on 73 CPUs on average.

Radial thinning is applied to particles reaching the observation level at distances < 200 m to the shower axis. This does not affect the evolution of the shower as it was applied after the particles reach the observation level, but it greatly reduces the disk space needed to store the shower. In total, the shower takes up 730 GB of disk space. For convenient analysis, the particles were re-sorted in 1008 files, corresponding to $42 \text{ radial} \times 24 \text{ azimuthal}$ bins.

# MW	ε_{em}	$w_{\text{max,em}}$	ε_{had}	$w_{\text{max,had}}$
0	10^{-16}	1	10^{-16}	1
1	10^{-8}	10^2	10^{-8}	1
2	10^{-7}	10^3	10^{-7}	10
3	10^{-6}	10^4	10^{-6}	10^2
4	10^{-5}	10^5	10^{-5}	10^3
5	10^{-4}	10^6	10^{-4}	10^4
6	10^{-8}	10^2	10^{-8}	10^2
7	10^{-7}	10^3	10^{-7}	10^3
8	10^{-6}	10^4	10^{-6}	10^4
9	10^{-6}	10^4	10^{-8}	10^4
10	10^{-6}	∞	10^{-6}	∞

TABLE II
Parameters of the different thinning methods applied in parallel. The thinning levels for electromagnetic and hadronic particles are given by ε_{em} and ε_{had} . The maximum weights for electromagnetic and hadronic particles are given by $w_{\text{max,em}}$ and $w_{\text{max,had}}$.

IV. ANALYSIS

What is the influence of thinning on observables at ground-level? The above mentioned azimuthal symmetry of a vertical shower permits the assumption that the observables depend primarily on the distance from the shower axis. The observation plane is divided in bins of 10 m^2 , a size close to the active area of the detector elements used in the Pierre Auger Observatory. The bins are arranged in circles around the shower axis at fixed distance r . At core distance r , the average for an arbitrary quantity p , \bar{p}_r , is formed from $p_{r,\varphi}$ over N bins (i) in azimuth φ . The fluctuations at r are given by the variance σ_r^2 .

$$\bar{p}_r = \frac{1}{N} \sum_{i=1}^N p_{r,\varphi} \quad \sigma_r^2 = \frac{1}{N-1} \sum_{i=1}^N (p_{r,\varphi} - \bar{p}_r)^2$$

The unthinning method [3] is used. The unthinning area is defined by $\Delta r/r = 0.1$ and $\Delta\varphi = 0.15$ radians. Only particles above the Cherenkov threshold in water are considered.

A. Particle Numbers

The quantities directly affected by the thinning and unthinning procedures are the numbers of particles of a certain type arriving in a bin (detector element). In the absence of substructure in the unthinned shower, one expects Poissonian fluctuations in the particle number between bins. However, in the thinned showers, one finds larger-than-Poissonian fluctuations after unthinning. This is because the Poissonian fluctuations on the smaller number of *tracked* particles are relatively larger. The increase of fluctuations is quantified by comparing the observed σ_r with the Poissonian fluctuations of the unthinned shower $\sigma_{r,\text{Poisson}}$ in the equivalent bin.

In figure 1 the ratio of $\sigma_r/\sigma_{r,\text{Poisson}}$ is shown for photons, electrons and positrons and muons, respectively, for the unthinned shower and different thinning levels. Without thinning, the electromagnetic particles show a deviation of few percent from Poissonian behaviour, possibly due to substructure in the shower. The fluctuations

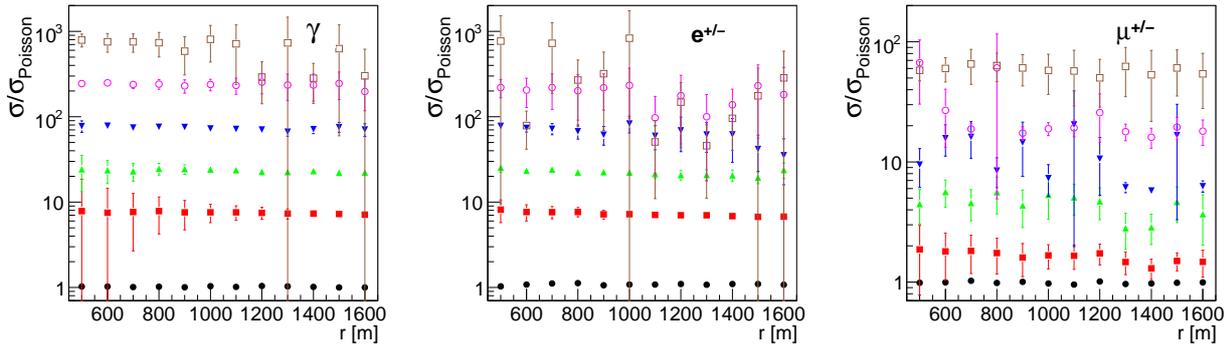


Fig. 1. The ratio of σ_r and $\sigma_{r,Poisson}$ for different thinning levels with optimal weight limitation. The markers correspond to : \bullet No thinning; \blacksquare 10^{-8} thinning; \blacktriangle 10^{-7} ; \blacktriangledown 10^{-6} ; \circ 10^{-5} ; \square 10^{-4}

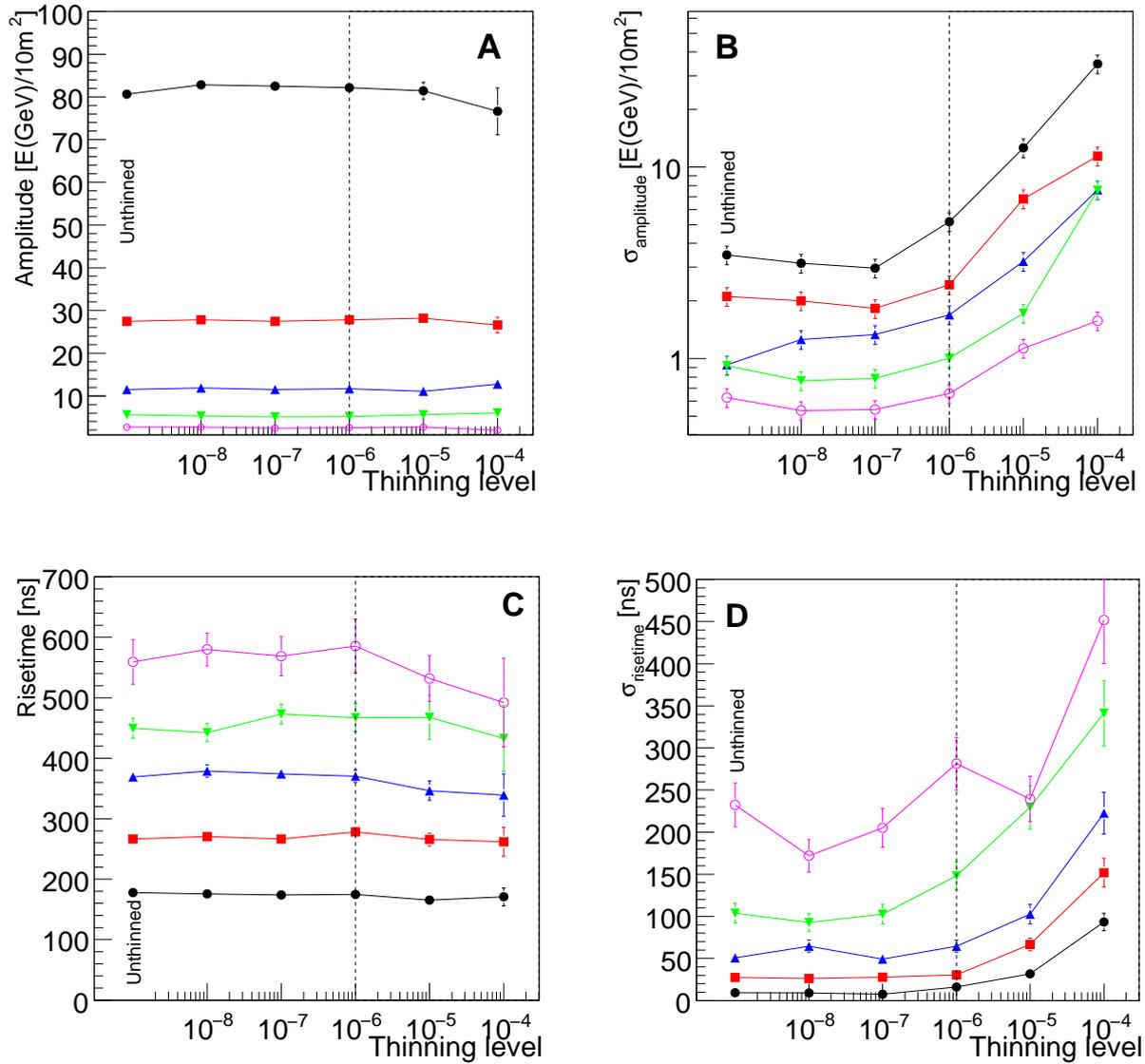


Fig. 2. This plot shows the amplitude, risetime and their standard deviations as function of thinning level for different core distances. The left most points are the values for the unthinned shower. The dashed vertical line indicates a commonly used value for the thinning level. The markers indicate the core distances: \bullet 600m; \blacksquare 800m; \blacktriangle 1000m; \blacktriangledown 1200m; \circ 1400m

in the muon numbers are Poissonian. When thinning is applied, the fluctuations become larger than Poissonian, as expected. The standard deviations scale with the thinning level. A factor of 10 increase in thinning level results in an approximately $\sqrt{10}$ increase in σ_r , due to the fact that a $10\times$ higher thinning level, decreases the number of tracked particles approximately by the same factor.

B. Observables

Air-shower experiments typically measure the energy the particles deposit in the active elements. In this study, it is assumed that the active volume is made up out of water, such as used in Haverah Park [5] and Pierre Auger [6]. Particles above the Cherenkov threshold emit light when passing through the water. The light is detected by photo-multipliers. A simple approximation is made to estimate the signal that could be measured in such an instrument. The signals generated by photons, electrons and positrons are proportional to the energy released above the Cherenkov threshold for electrons. Muons can maximally lose 0.24 GeV.

1) *Amplitude*: The amplitude is the total integrated measured signal in an active detector element. This quantity is typically used to reconstruct the core position of the shower and the energy of the primary particle [7]. Fig. 2 a) and b) show the amplitude and its fluctuations as function of thinning level for several core distances. The amplitude remains constant when the shower is thinned at increasingly higher levels, i.e. no significant bias is introduced due to the thinning and unthinning procedures. The standard deviation of the amplitude stays relatively constant up to a thinning level of 10^{-6} , but then increases significantly for all distances. The effect weakens with increasing core distance, as closer

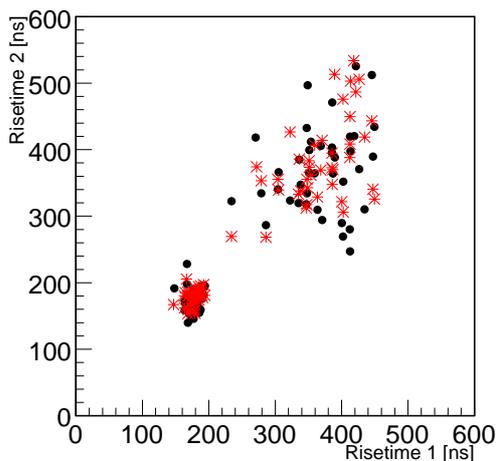


Fig. 3. Fluctuations due to thinning and sampling. Horizontal axis: true risetime from the unthinned shower. Vertical axis for black dots: risetime as calculated with 10^{-6} thinning. Vertical axis for red stars: risetime in a directly adjacent detector element without thinning. Each point represents one of 40 detector elements placed on a circle at a fixed core distance. The two different clouds of points correspond to core distances of 600 m (lower) and 1000 m (upper).

to the shower core the highest weights are found.

2) *Risetime*: The risetime is the time it takes for the integrated signal to grow from 10% to 50% of the total signal in a detector element. It is related to the longitudinal development of the shower and can be used to extract information on the mass of the primary particle [8]. Fig. 2 c) and d) show the risetime and its fluctuations as function of thinning level for several core distances. As with the amplitude, no bias is introduced due to the thinning over most of the range, as the dependence on thinning level is almost flat. The standard deviation of the risetime increases significantly with increasing thinning level from 10^{-6} . Thus, up to a thinning level of 10^{-6} , the fluctuations between detector elements (sampling fluctuations) are larger than those introduced by the thinning. This large scatter between individual detector elements is shown again in figure 3. The scatter of the risetime introduced by 10^{-6} thinning (black points) is of the same order (~ 70 ns at 1000 m) as the scatter in an unthinned shower between adjacent detector elements (red stars).

At 1000 m the risetime uncertainty is of the order of ± 75 ns, which will make it difficult to separate proton from iron induced showers on the basis of risetime at 10^{19} eV. The expected risetime differences between p and Fe at 10^{19} eV are only about 30 ns. For thinning levels better than 10^{-6} the situation will not improve as then the sampling uncertainties due to the 10 m^2 detector elements dominate.

V. CONCLUSIONS

A 10^{19} eV proton extensive air shower has been fully simulated without the use of particle thinning techniques. This has been done using an extended version of CORSIKA which not only parallelises the calculation but also provides the possibility of studying the effects of statistical thinning by calculating unthinned and thinned showers simultaneously. These features will be available in future distributions of CORSIKA. The thinning procedures best suited for specific high-energy cosmic ray instruments depend on the size and arrangement of their detectors. Simulations should be so detailed that thinning artefacts remain smaller than uncertainties due to sampling and detector effects. With CORSIKA a tool is available that allows investigation of uncertainties in great detail.

REFERENCES

- [1] A. M. Hillas, Proc. 17th ICRC, Paris 8 (1981) 193
- [2] M. Kobal, Astropart. Phys. **15** (2001) 259.
- [3] P. Billoir, Astropart. Phys. **30** (2008) 270.
- [4] D. Heck et al., Report FZKA 6019 (1998), Forschungszentrum Karlsruhe
- [5] M.A. Lawrence, R.J.O. Reid and A.A. Watson, J. Phys. G: Nucl. Part. Phys. **17** (1991) 733
- [6] J. Abraham *et al.* [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **523** (2004) 50.
- [7] J. Abraham *et al.* [Pierre Auger Collaboration], Phys. Rev. Lett. **101** (2008) 061101 [arXiv:0806.4302 [astro-ph]].
- [8] J. Abraham *et al.* [Pierre Auger Collaboration], Astropart. Phys. **29** (2008) 243 [arXiv:0712.1147 [astro-ph]].