Auger project technical note — A case for MOCCA shower simulation runs at 10^{-7} thinning level

Clement Pryke
Enrico Fermi Institute, University of Chicago, USA

January 2, 1999

Abstract

The current predictions of the Auger ground array performance are based on detailed simulation of the water Čerenkov detector performance followed by reconstruction of the simulated event data so derived. However, until this point the input information to this process has represented the mean of 100 shower simulation runs; shower-to-shower fluctuations resulting from *physical* effects have been sacrificed in order to reduce the *statistical* artifacts which result from the thinning algorithm. In this note it is argued that reducing the thinning threshold by one order of magnitude to 10^{-7} of primary energy will provide sufficiently high statistics that the summarized results from single showers can be input into the ground array simulation process. At this thinning level the predicted experimental deviation from ideal performance will be dominated by the instrumental limitations of the proposed Auger arrays, although a component will remain from the residual computational limitations in shower modeling.

1 Introduction

The Auger Design Report document calls for two surface arrays of $\approx 1000 \times 10$ m^2 water Čerenkov detector units arranged in a hexagonal grid with inter-detector spacing of 1.5 km [1]. In order to predict the experimental performance of such a configuration detailed modeling of Extensive Air Shower

(EAS) development is required, together with modeling of the interaction of the shower front particles with the detector units themselves.

Up to this time the MOCCA program written by Michael Hillas has been used to provide the shower simulation function. MOCCA is a detailed 4-dimensional cascade including a realistic atmospheric model, and implementing a technique known as thin-sampling to allow the generation of extreme high energy showers with reduced computation time. However the use of "thinning" imposes limitations for the present application which are the subject of this note.

Custom code was written during the Fermilab Design Workshop in the first half of 1995 to perform the detector simulation: this has now been dubbed AGAsim for Auger Ground Array simulation. The techniques used in AGAsim are described in detail elsewhere [2] and will not be repeated here. Basically the program uses the binned and summarized output available from MOCCA to generate complete lists of shower particles incident on the detector units which take part in a given cosmic ray event. Each particle is then tracked through each detector, interacting and radiating Čerenkov photons which are then ray-traced until absorbed or incident on a photomultiplier tube. This is a full "brute-force" approach which can be expected to reproduce nearly all instrumental limitations which a real experimental system will be subject to.

The final stage of AGAsim is the reconstruction of the simulated shower events in the same manner as would be performed using real data, and comparison with the known initial event parameters to make a prediction of the experimental performance. Figure 6.12 in the Design Report represents the results of early reconstruction algorithms implemented by Huynh Phuong [3]; improved results are now available from a new reconstruction code — these will be mentioned below, and detailed in a future Auger technical note.

AGAsim relies on results from MOCCA to generate its lists of incident particles at each detector. This information is provided in the form of a "summary file" which is a large 4-dimensional array recording particle density broken down by type, perpendicular core distance, arrival time delay (relative to the tangential shower front plane), and energy. In this and previous work there are 3 bins of particle type (γ, $e^{\pm}$ and $\mu^{\pm}$), 20 logarithmic bins of core distance (5 per decade from 50 m to 5 km), 30 logarithmic bins of arrival time (10 per decade from 10 ns to 10 μs), and 40 logarithmic bins of particle energy (0.1 MeV to 10 TeV)[1].

2 The limitations of 10^{-6} thinned shower simulation runs

The "thinning level" used in MOCCA is conveniently expressed in terms of a fraction of the primary particle energy. Above this energy all shower particles are followed; below it only a fraction are tracked, and each of these is assigned a weighting factor which exactly accounts for the energy disguarded in unfollowed particles [4]. The most basic MOCCA output format is a simple list of ground level particles giving type, weight, impact coordinates, energy, arrival time and incidence angle.

All AGAsim running reported before now (including that whose results are described in the Design Report) made use of MOCCA summary files as input that where extracted from batches of 100 1×10^{19} eV showers thinned at 10^{-6} of primary energy (10 TeV). This was believed to be necessary to overcome the artificial shower-to-shower fluctuations introduced by the thin sampling algorithm. However, in the process any such fluctuations whose origin is physical will also be averaged out. Hence the predicted deviation from ideal experimental behavior will represent the instrumental limitations and can only be regarded as a "best-case" scenario.

It is well known that there are significant shower-to-shower fluctuations in the real world. These are caused by fluctuations in the first interaction point of the primary particle and in the stochastic nature of the initial ultra high energy cascade generations due to the small number of particles involved.

A vertical 10^{19} eV MOCCA shower thinned at 10^{-6} produces a ground particle list with $\approx 50,000$ entries. This seems like a large number; however, the vast bulk of the particles are located close to the shower axis. Due to the wide spacing of the Auger arrays (1.5 km) there will only be one detector at < 650 m from core in any given (near vertical) event. Even in the rare instance that a shower core does fall very close to a detector unit it is unlikely that it will record useful information — it will not be

[1] Bin upper edge values are given by $r_u = 0.05 \times 10^{i/10}$, where $1 \leq i \leq 20$ and r_u is in meters, $t_u = 10 \times 10^{i/10}$, where $1 \leq i \leq 30$ and t_u is in ns, and $E_u = 0.1 \times 10^{i/5}$, where $1 \leq i \leq 40$ and E_u is in MeV. Note that particles arriving at $t < 10$ ns are placed in the first time bin.

worth the expense required to extend the dynamic range of the detectors by the enormous margin necessary to obtain data in this small sub-set of instances. Hence for practical reasons we need to consider the experimentally measurable EAS signal at core distances of ≥ 500 m.

The top left plot in figure 2 illustrates the lack of (electromagnetic) ground particle list entries at large core distances in a 10^{-6} thinned event; the superimposed "target pattern" shows the annular core distance bins mentioned in the previous section. At top right the familiar lateral distribution curves are shown; the raw entries have been weighted, binned against core distance and normalized according to the annular bin areas. A 10^-7 thinned event is shown for comparison in the lower plots: note that there are a large number of list entries at core distances as great as 2 km, and the lateral distribution curves are well behaved as far out as 3 km.

3 Impact of limited shower simulation statistics on predictions of experimental performance

What effect does this lack of ground particle statistics have on the predicted measurement power of the proposed Auger ground array systems? To try and answer this using the available library of 10^{19} eV 10^{-6} thinned MOCCA showers some special AGAsim runs were made. The most difficult and sensitive reconstructed event parameters are those related to the nature of the primary cosmic ray particle (proton, iron nucleus etc.). The new reconstruction algorithms mentioned in the introduction use two measurements of the "gross time structure" of the simulated water Čerenkov detector output signal. These are t_{10} — the time between the arrival of the reconstructed tangential shower plane and the 10% point of the total time integrated signal, and t_{50} — the time between the 10% point and the 50% point of the signal. For each simulated event a simple linear fit is made to the available t_{10} and t_{50} values versus core distance, and the gradient used as a composition sensitive parameter.

Figure 4 illustrates the reconstruction results from batches of 100 showers initiated by protons and iron nuclei. In the upper plot all 100 showers of each type were generated using a single MOCCA summary file, itself extracted from a full set of 100 showers thinned at 10^{-6}. As already mentioned this is the “mode” in which all previous running has been carried out. It is seen that there is a very good separation between the proton and iron showers. At lower left is an identical plot except that in this case 10 events were generated from each of 10 summary files, each of these being extracted from 10 of the 10^{-6} thinned showers. Note that the species separation is nearly as good as in the “mean of 100” case, and also that a 10 showers thinned at 10^{-6} is roughly equivalent in terms of ground particle statistics to a *single* 10^{-7} thinned run. The final plot at lower right shows the results obtained when one event was generated from each of 100 summary files, each of these being derived from a single 10^{-6} thinned shower. There is still primary type separation, although the impact of the limited ground particle statistics is clearly seen.

4 Practicalities of MOCCA running at 10^{-7} thinning level

It would clearly be ideal if we could generate many hundreds of 10^{-7} thinned MOCCA showers to be input individually into the Auger ground array simulation process. As shown in the previous section the residual departures from ideal experimental performance would then be dominated by the instrumental limitations of the proposed system, and “worst-case” predictions could be made including *nearly all detrimental effects*. Is it practical to contemplate such a computational effort?

So far I have successfully run five 10^{-7} thinned showers using a Silicon Graphics computer at Fermilab having 200 MHz R4400 processors. This machine has 20 CPUs, although normally we should not occupy more than 2 or 3 at a time. However there is the possibility that we can use ≥ 5 state-of-the-art UNIX Alpha boxes if we want to.

Dave Nitz and Jim Mathews at Michigan have provided a version of MOCCA which is essentially unchanged from the most recent code which the author Michael Hillas has seen fit to distribute. This is called MOCCA92t, and was interfaced to the Bartol hadronic interaction generator SIYBLL by Hillas and others at Bartol; there is a simple run time switch which allows SIYBLL to turned on or off. This so called “Michigan version” makes no physics changes from MOCCA92t, but does incorporate a system of complier specific tags and a filter algorithm (Lex script) which allows the same

Energy (eV)	θ (degrees)	Run time (hours)	ASCII file size (Mbyte)	Compressed file size (Mbyte)
1×10^{19}	0	24.0	36.7	10.4 (28%)
1×10^{19}	30	27.6	37.6	10.8 (29%)
1×10^{19}	45	30.5	26.4	7.7 (29%)
1×10^{19}	60	32.5	8.6	2.5 (29%)
1×10^{20}	30	34.5	35.9	10.4 (29%)

Table 1: Run parameters for 10^{-7} thinned MOCCA shower simulation runs.

platform independent source file to be compiled and run on several different types of computer (currently SGI UNIX, HP UNIX and Dec VMS are supported. This version of MOCCA *lacks* the binary ground particle output file format originally incorporated into an older version of MOCCA (MOCCA89) by Corbin Covault at Chicago, and used for the 10^{-6} thinned shower library runs.

In order to reduce the particle list file size when using ASCII output and running at 10^{-7} thinning level I have set the core distance cut radius within which a particle must fall to be included in the output file to $50 < r < 5000$ m. There are in any case few particles at larger radii, and those at smaller cannot be regarded as experimentally measurable for the reasons given in the previous section, and hence are not very interesting in the context of the current work. The run times and output file sizes for the test runs so far generated are given in table 1. The input parameters in each case were a proton primary of observed at an atmospheric depth[2] of 850 g cm^{-2}. The optional SIYBLL hadronic interaction generator was not used in these runs. Since the output files are ASCII they can be stored in compressed format when not in use; I also give the file sizes when reduced with the standard UNIX "compress" utility.

Figure 3 is equivalent to figure 2 but shows the results for the 30° and 45° 10^{-7} thinned showers. Note that there is significant evidence of azimuthal asymmetry in the shower "footprint". This is a cause for concern as the current annular binning strategy will average this out, and hence unrealistically

[2] I have switched to an atmospheric depth of 850 g cm^{-2} as opposed to 1000 g cm^{-2} as used in the bulk of previous running as this is closer to the altitude of the favored Argentinian site, and probable Northern Hemisphere sites.

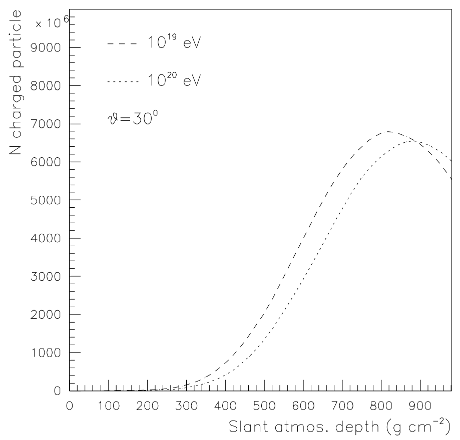

Figure 1: Longitudinal shower profiles from 10^{-7} thinned proton initiated showers of 10^{19} eV and 10^{20} eV. The vertical scale should be multiplied by 10 when reading the 10^{20} eV event.

good results may be obtained from the event reconstruction studies. On the other hand reconstruction algorithms can certainly be developed which attempt to take account of the expected asymmetry — further investigation is called for.

Longitudinal shower profiles are also available from these runs; see figure 1. The longitudinal profiles of the $30°$ 10^{-7} thinned showers at 10^{19} eV and 10^{20} eV are plotted. The atmospheric depth is shown along the slant; for this zenith 981 g cm^{-2} corresponds to ground level at the proposed site altitude.

5 Conclusions and plans

I believe that it has been demonstrated that shower simulation runs are both possible and desirable at a thinning level of 10^{-7} of primary energy. This would allow (definitive?) predictions of the Auger ground array performance taking into account all shower-to-shower fluctuations and instrumental limitations. However it would represent a huge compute effort and most likely the use of resources in several parts of the world. With the price of modern disk storage capacity, and making a simple cut on particle core distance for output there is no pressing need to use a binary particle list format; ASCII files can in any case be compressed when not in use, and for transmission. However there is certainly no strong reason not to use such a format. The summary files are very small; ≈ 25 Kbyte when compressed.

I suggest that the MOCCA input files (also called run-cards) for all these runs be generated at a single site and distributed in batches to be run elsewhere. The MOCCA code used should also be maintained at a central site; the existing Michigan version represents a good model of a platform independent source file which can be compiled and run on several architectures and operating systems. Unless there are strong arguments to the contrary I will start work using the Fermilab machines and the Michigan code more or less immediately. However, the Slovenian group have access to an extremely powerful machine (64 CPU Convex SPP1000) which is not heavily loaded at present so it is probably most important what they choose to do. The Australian group (Adelaide) have also been running 10^{-7} thinned showers (using the Michigan code).

I would like to open for debate the following possibility: should we generate showers sampled from a spectrum of energies and zenith angles rather than the approach which has been followed thus far of running batches at several discrete energies and angles? I would propose generating showers from say 10^{19} eV to $10^{20.5}$ eV, with zenith angles from $0°$ to $65°$. Zenith angles would be sampled from a distribution which goes as $\cos\theta \sin\theta$ (the number of events with cores falling within a given area of ground). The energy distribution is more tricky; if we use the $\gamma \approx 2$ power law spectral form which is observed in nature we will have too few events at the high energy end in which we are primarily interested. Perhaps we could arbitrarily set $\gamma = 1$? Also including a deliberate kink in the energy spectrum might be very interesting.

All of the above represents my own (current!) opinions only. I urge anyone who disagrees with anything I have said to get in touch (email pryke@hep.uchicago.edu).

References

[1] The Auger Collaboration, *Design Report* (October 1995) Fermi National Accelerator Laboratory.

[2] C.L. Pryke, PhD thesis, *Instrumentation development and experimental design for a next generation detector of the highest energy cosmic rays* (The University of Leeds 1996).

[3] H. Phuong, *Shower Simulation Report,* Auger proj. tech. note GAP-95-026, (October 1995) Fermi National Accelerator Laboratory.

[4] A.M. Hillas, *Proceedings of the 17th International Cosmic Ray Conference (Paris),* **8** (1981) 193.

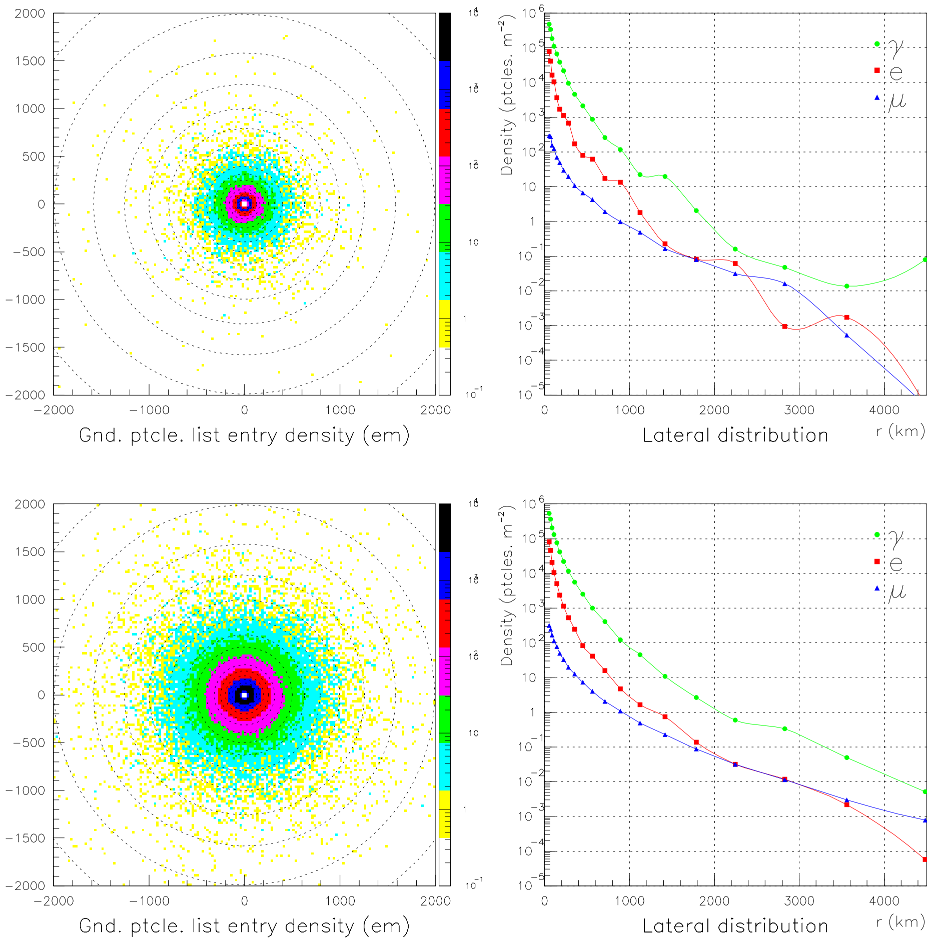

Figure 2: Ground particle results from MOCCA running at thinning levels of 10^{-6} — upper plots, and 10^{-7} — lower plots. The event parameters were vertical primary proton of 10^{19} eV energy observed at 850 g cm^{-2} atmospheric depth in both cases. The left side shows the density of output list entries (electromagnetic particles only) in the ground plane; the "target" pattern indicates the annular bins used when summarizing the data. The right plots show the lateral distribution of particles from the same runs; the numerical density of weighted ground particles is shown (in this case for all particles with $E > 0.2$ MeV).

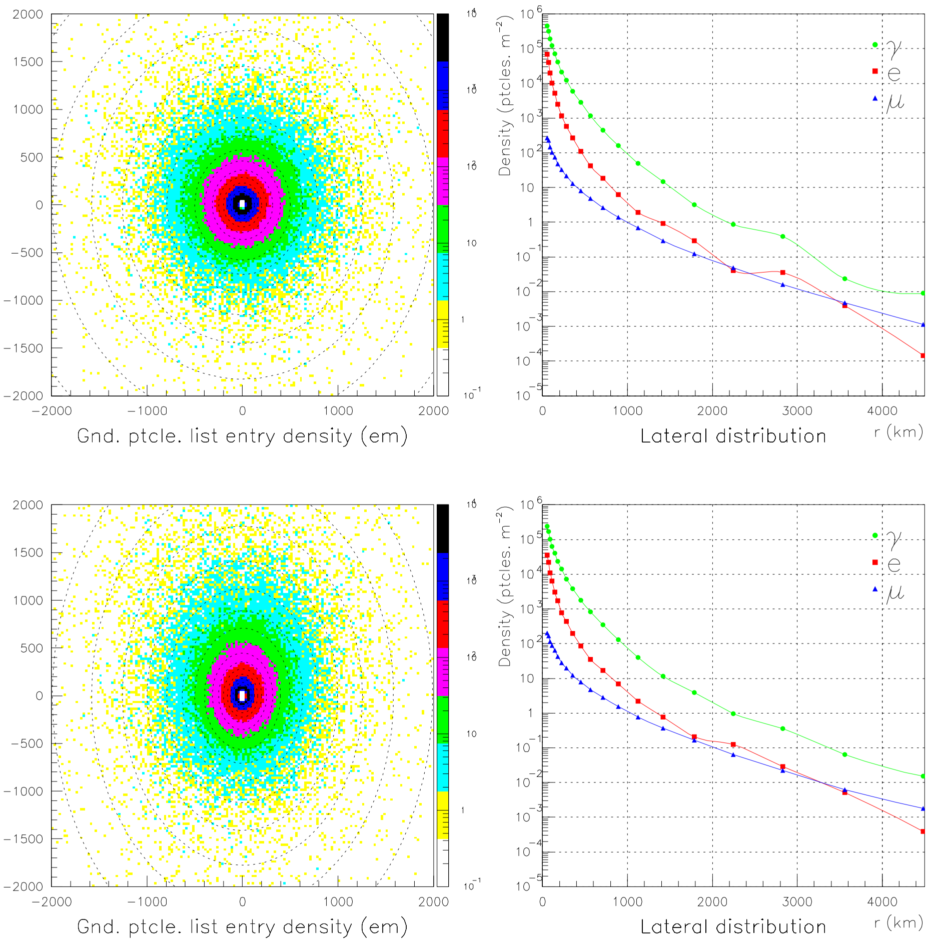

Figure 3: This figure is equivalent to figure 2. The distribution of electromagnetic ground particle entries and lateral distributions are shown for single 10^{-7} thinned 10^{19} eV proton showers observed at 850 g cm^{-2} atmospheric depth. The upper plots are for a shower at zenith angle of 30° incident from the north (ie down the page), the lower for a 45° event.

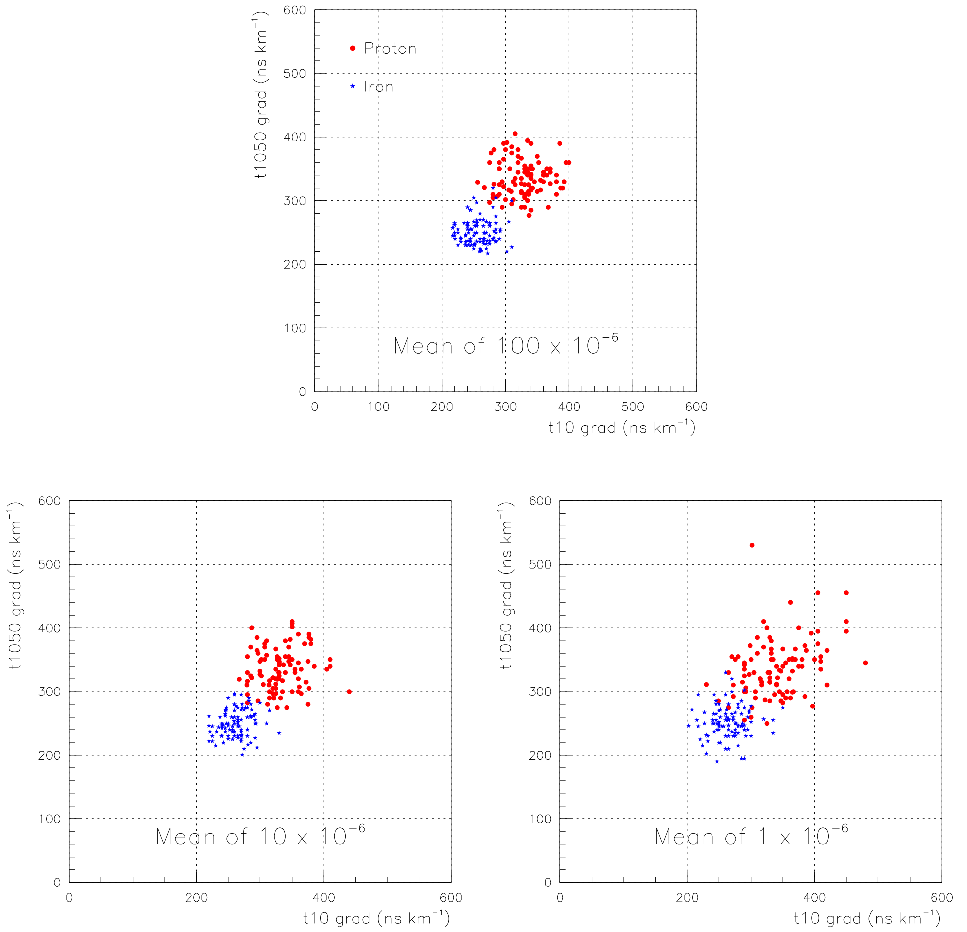

Figure 4: Results from reconstruction of simulated Auger ground array data. Each plot is a scatter diagram of two shower front time structure parameters which exhibit compositional sensitivity. At top the MOCCA output summary file used as input to the ground array simulation process was extracted from a set of 100 showers thinned at 10^{-6}. At lower left summaries of 10 showers were used, and at right single showers. (Event parameters were $E = 40$ EeV, $\theta = 30°$, in all cases.)