Challenges of simulating air showers at ultra-high energy

Dieter Heck, Ralph Engel, Tanguy Pierog,
Institute for Nuclear Physics, Karlsruhe Institute of Technology

Gurpratap Thiara, Shenan Karla, Himani Singla,
PEC University of Technology, Chandigarh, India

Gevorg Poghosyan, David Seldner, Frank Schmitz
Steinbuch Centre for Computing, Karlsruhe Institute of Technology

Abstract

In this contribution we describe the work on the parallelization of the Monte Carlo code CORSIKA to master the challenges of simulations for development of an Extensive Air Shower initiated by cosmic ray particles at the highest energies of 10^{20} eV. We have implemented workflow system for parallel execution of CORSIKA on distributed computing infrastructures and make preliminary tests simulations for energies up to 10^{18} eV. It allows us to estimate computational needs and challenges for productive simulations at higher energies to escape possible overflows when using parallel job scheduling systems.

Introduction

One of the problems in simulating Extensive Air Showers (EAS) initiated by cosmic rays at the highest observed energies comes from the huge number of 10^{11} particles which are created within the avalanche by the consecutive interactions with air. Their fate is determined by the flight path between their creation and their interaction or decay, and their nature and energy determines the number, types, energies and directions of the secondary particles generated in the interactions. These stochastic processes are well modelled by Monte Carlo (MC) methods to get correctly not only the mean values (which might be given eventually by analytical approximations) but also the fluctuations and correlations of the measurable shower observables as e.g. lateral distribution of the particles at ground or energy spectra of those particles at different distances to the shower axis. A solution of following this huge number of particles might be the parallelization of the MC code CORSIKA (COsmic Ray SImulations for KAscade) (Heck et al. 1998) to treat independent sub-showers by different cores in parallel. Already in an early stage of the development of the CORSIKA code several possibilities to circumvent the long CPU times which increase nearly linearly with increasing energy of the initiating cosmic particle, were considered (Gils et al. 1993) including various structures of a multi-Transputer computer farm (Gils et al. 1989).

Today the availability of multi-core processor systems has revived the parallelization ideas. To master challenges of productive usage of current and future high-end multi-core processors systems an innovative research and support structure, so-called Simulation Laboratories (SimLab), are currently being established at different supercomputing centres in Germany. In particular “SimLab Astro- & Elementary Particle Physics” (SimLab A&E Particle) at SCC of KIT is providing support on parallelization of scientific application CORSIKA and porting it into up-to-date distributed and super-computing infrastructures in Europe (Thiara, Poghosyan und Schmitz 2009).

Outline of Problem

Ultra-high energy Cosmic Rays

The highest energies of cosmic rays observed so far exceed 10^{20} eV. To accelerate protons to those energies with the present technologies of the LHC-collider, one needs an accelerator ring with the diameter of Mercury's orbit around the Sun and an acceleration time of more than 800 years. Nevertheless such Ultra High Energy (UHE) particles bombard the Earth's atmosphere from space and the resulting shower cascades are observed in experiments. To interpret these observations correctly simulation software is necessary which can model such events.

"CORSIKA" simulation software

During the past 20 years the code package CORSIKA (Heck et al. 1998) for the simulation of EAS has been developed at KIT and became a worldwide standard instrument for scientists, studying the complex simulation of air-showers induced by cosmic rays in the atmosphere. CORSIKA is based on the MC technique and allows the realistic simulation of interaction, propagation, and decay of particles in extensive particle cascades initiated by high-energy cosmic rays from our Galaxy or extragalactic objects. CORSIKA has become the worldwide standard tool for EAS simulations. More than 750 users apply this code in about 40 experiments worldwide.

One method to reduce the CPU-time is the so called "thinning" (Kobal 2001). In this method after a certain stage of the shower development only statistically selected particles are followed while the bulk of particles are dropped. The particles followed get a weight to keep the energy balance within the EAS. The drawback is additional artificial fluctuations of observable shower quantities, and it is highly desirable to verify that the results are not biased by the thinning procedure in any respect. This needs the verification by fully simulated EAS without thinning even at UHEs.

But simulation for UHECR with energies as high as 10^{20} eV with the present version of CORSIKA applicable for sequential runs only, could take decades on single computers (Knapp et al. 2003, Schmidt and Knapp 2005). Hence, implementation of a new methodology for parallelization is vital to manage simulation for such UHECR at mentioned energies in acceptable time.

Scientific Workflow for Parallel executions of "CORSIKA"

First runs of parallel CORSIKA executions were performed on the Karlsruhe multi-Transputer system (Gils et al. 1989) simulating 10000 showers with up to 10^{15}eV for energies. The limitation for output was set by the size of the storage discs, at that time (1989) 500 MByte.

Currently we developed scientific workflow management systems for parallel; generation-type execution of CORSIKA simulated air showers. It allows splitting and distributing input data for corresponding sub-showers to start individual parallel jobs per secondary particle on own node/computer/core. Considering simulation of separate sub-showers as new initial simulations, the parallel runs are fully independent and simulation easily scalable on any distributed computational system. Different amount of sub-showers could be generated by varying the input parameters for energy gap. Secondary particles energy of which is lying in the given gap are grouped or separately simulated.

Implementing initial book keeping system, we limit cascade-like structures into generations to balance computation needs of simulations on computational infrastructure, where the simulation must run. It could be adapted and optimised, for instance to escape long jobs waiting in queues or overflowing the scheduling system.

Hitherto, we performed test runs using LoadLeveler driven computing infrastructure of SCC called OpusCluster. We have simulated proton induced showers with initial energies 10, 100 and 1000 Peta electron Volt (PeV=10^{15}eV), when higher limit of energy gap to start separation of secondary shower simulations are varied from 0.01 PeV till initial energy. For particle at energies lower than "ground" limit of gap equal to 0.001 PeV, no any separate simulations started.

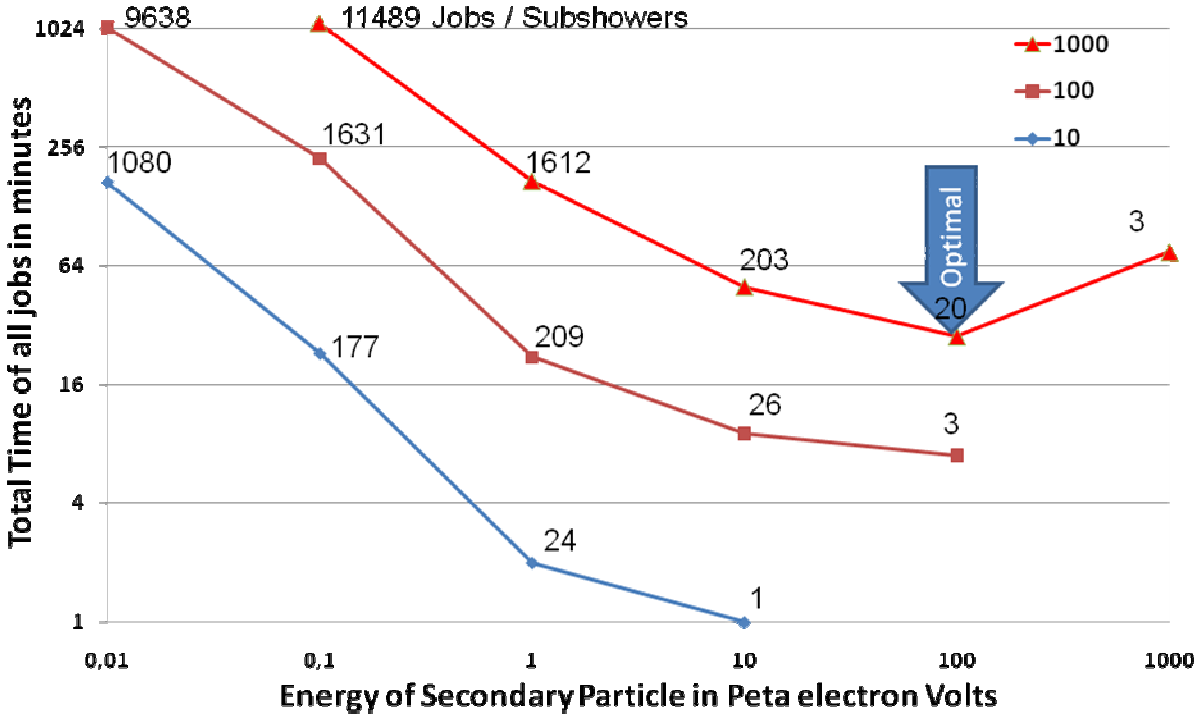

Fig. 1 Total time spent for simulation of proton induced showers with different initial energies and energy gaps. Numbers on graph, show amount of jobs/subshowers. It could be seen that parallel simulations could be optimally used only starting at high energies and enough subshowers.

In present system, each parallel calculation stores results into local or shared disk space accessible from particular node, where it is running. Here the variation of parameters allows escaping limitations on storage space available on particular node/system, when more sub-showers with smaller outputs would be used. Hereafter, automated tracking and book keeping system must be improved as of large amount of data up to 10^6 files with total size of up to 200GB per EAS is expected to be generated during full simulation. The last must be analyzed and compared with experimental results in "real-time".

Challenges – Work in progress

It is indispensable to implement modern data recording and analyzing algorithms into the developed Workflow system or CORSIKA code itself. Parallel runs and possible generation of hundred of thousand small files will consequence of troubles in file system, when even simply trying to read, analyse such amount of files. In addition, the problem of uncompleted or improperly simulated sub-showers could raise the complexity of management for generated big data amount, as of revealed fault sub-showers and corresponding resulting files must be simulated again. Here a algorithms of identification of sharing and re-running identified sub-showers will be improved.

Subsequently optimization and internal parallelization in the CORSIKA source-code itself will be next attempted with goal to reach maximal scalability of code on distributed and supercomputing systems.

References

Gils H.J., Heck D., Oehlschläger J., Schatz G. and Thouw T., Computer Physics Communications 56 (1989) 105

Gils J. , Heck D., Oehlschläger J., Schatz G. and Thouw T., *A Multi-Transputer System for Parallel Monte Carlo Simulations of Extensive Air Showers*, in: Report KfK 5185 (1993), ed. W. Hohenhinnebusch

Heck D., Knapp J., Capdevielle J.N., Schatz G., and Thouw T.. *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers.* 90 pages, 10 figures: Forschungszentrum Karlsruhe Report FZKA 6019 (1998)

Knapp J. et al., Astropart. Phys. 19 (2003) 77

Kobal M., Astropart. Phys. 15 (2001) 259

Schmidt F. and Knapp J.: *Results from the simulation of an unthinned proton shower at* $5x10^{18}$ *eV*, GAP note 2005-095 (www.auger.org/admin/GAP_Notes/)

Thiara G., Poghosyan G. and Schmitz F. "*SCC works on development of different scientific simulation codes*." SCC-NEWS, no. 2 (2009)