Validation of the Geant4 Bertini Cascade model and data analysis using the Parallel ROOT Facility on a Linux cluster

A. Heikkinen, T. Lindén, Helsinki Insitute of Physics, PO Box 64, FIN-00014 Helsinki, Finland

Abstract

Validation of hadronic physics processes of the Geant4 simulation toolkit is a very important task to ensure adequate physics results for the experiments being built at the Large Hadron Collider. We report on simulation results obtained using the Geant4 Bertini cascade isotope production cross-sections for Pb and Au target materials and incident proton kinetic energies between 1–2.5 GeV.

The cross-section benchmark study in this work has been performed using a Linux cluster set up with the Red Hat Linux based NPACI Rocks Cluster Distribution. For analysis of the validation data we have used the Parallel ROOT Facility (PROOF). PROOF has been designed for setting up a parallel data analysis environment in an inhomogeneous computing environment. Here we use a homogeneous Rocks cluster and automatic class generation for PROOF event data-analysis.

INTRODUCTION

The Bertini cascade model [1] is implemented as part of the Geant4 hadronic shower framework [2]. It contains an intra-nuclear cascade model with exitons, as well as a pre-equilibrium model, a nucleus explosion model, a fission model, and a evaporation model. The Bertini cascade models may be used to simulate pion-, proton-, and neutron-induced reactions in nuclei. It is valid for incident energies up to 10 GeV, making it useful for the simulation of hadronic calorimeters. The potential of the models for instrumentation in HEP, space and medicine has been recently evaluated by V. N. Ivanchenko in [3] and [4]. In nuclear physics intra-nuclear cascade codes have been successfully used to design Accelerator Driven Systems to simulate spallation targets [5].

BaBar has been the first major experiment to evaluate the Geant4 Bertini cascade code [6]. In comparison to the old Geant3 GEISHA-based physics, Geant4 Bertini cascade performs favorably. Especially pion production is superior compared to parametrized Geant3 models. To provide a more detailed treatment of hadronic calorimetry, and kaon interactions in general, the Bertini model is being extended to include incident kaons up to an energy of 15 GeV. D. Wright reported at CHEP04 on adding kaons to the Bertini cascade model [7].

BERTINI CASCADE VALIDATION WITH PROOF

Most of the validations of the Geant4 Bertini cascade model so far have been concentrated below 1 GeV incident bullet energies and focus has been on double differential hadron cross sections. Large Hadron Collider Computing Grid (LCG) validation results of the Bertini cascade model can be found from [8].

Only a few papers have been published on validation of the Geant4 Bertini cascade isotope production. The residual production yield of Geant4 Bertini cascade models has been studied previously by Y. E. Titarenko *et al.* in [9] for 0.1, 0.2, 0.8 and 2.6 GeV proton irradiated ^{nat}Hg targets. They compared the INUCL code, which is an original fortran version of the Bertini cascade models discussed here against the LAHET, CEM95, CEM2k, CASCADE, and YIELDX codes.

In line with systematic validation of Geant4 hadronic physics for large HEP detectors [10, 11, 12], we present here a validation program for Bertini cascade isotope production for incident proton energies $E \geq 1$ GeV. We have chosen some representative cases from the literature to be studied in detail. The following table summarizes our validation program.

Table 1: A summary of our validation program.

Proton energy [GeV]	*Target*	*Reference*
1.0	^{208}Pb	[5, 13]
1.4	^{208}Pb	[13]
1.5	^{208}Pb	[14, 15]
1.8	^{197}Au	[16]
2.5	^{197}Au	[17, 18]

We have used Geant4 6.2 with LCG compliant Bertini cascade PACK 2.4 and ROOT 4.00/08. Several ROOT database-files were created for each configuration of 200,000 events. We created a ROOT `TTree` for the collision events and utilized the `TSelector` framework together with PROOF data collection `TDSet`, as demonstrated by the ROOT team in CHEP03 conference [19, 20]. A skeleton for the analysis code was generated with `TTree::MakeSelector()` function.

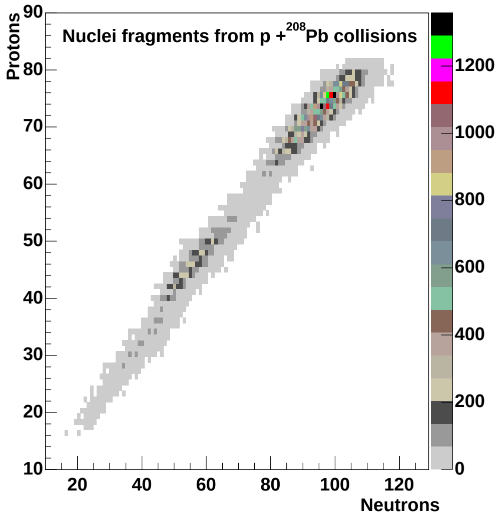

Figure 1: Isotope production from p(1 GeV) + 208Pb collisions with Geant4 Bertini intra-nuclear cascade models.

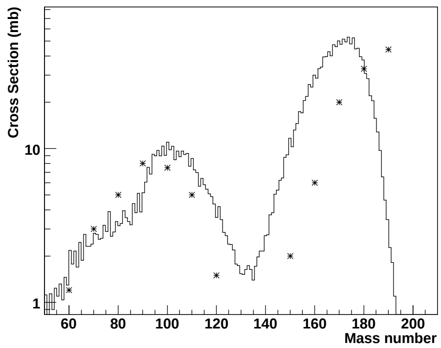

Figure 2: Geant4 Bertini cascade model simulation of fragment mass spectrum in the p(1GeV) + 208Pb reaction. Data points are from [5]. The fission peak centered around $A \sim 90$ is reproduced quite well. Apparent difference is seen in the fragmentation peak, which collects heavy residues after evaporation of the remnant.

ISOTOPE PRODUCTION CROSS-SECTIONS RESULTS

With the help of ROOT and PROOF we have compared simulation results of the Geant4 Bertini cascade with experimental data. When compared to the literature, the general performance of Bertini cascade seems to be relatively good. We compared experimental production cross sections by M. Enke *et al.* in [16] with p(1.8 GeV) + ^{197}Au simulations. Figure 3 shows an example of these simulation results. In comparison to [16] a good agreement is seen. Comparison of experimental data in [5, 13] with Figures 1 and 2 relieve some discrepancies, which are partly due to the different experimental setup.

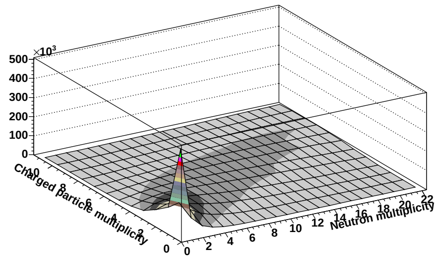

Figure 3: Correlation of charged particles versus neutron multiplicity in the p(1.8 GeV)+ ^{197}Au reaction.

NPACI ROCKS

NPACI Rocks Cluster Distribution is a RPM based cluster management software for scientific computation based on Red Hat Linux. The latest versions are based on a recompiled Red Hat Enterprise Linux v. 3.0. Version 3.0.0 used in this work with kernel 2.4.20-20.7 is still based on Red Hat 7.3. There are at least five Rocks based clusters on the June 2004 Top500 list, at positions 58, 281, 286, 413 and 430. There are some 300 registered Rocks clusters with more than 16800 CPUs with an theoretical estimated peak of 65 TFLOPS. In comparison the largest supercomputer on the Top500 list, The Earth Simulator has a theoretical peak processing power of 41 TFLOPS and the SETI@HOME project running on a million computers is estimated to have a computing power of 60 TFLOPS.

All Rocks nodes are considered to have soft state and any upgrade, installation or configuration change is done by node reinstallation, which takes about 6–15 min for a node and about 20–30 min for a whole cluster depending on the speed of the nodes.

The default configuration is to reinstall a node also after each power down. Settings like this can be easily changed according to taste. Rocks makes it possible for nonexperts to setup a Linux cluster for scientific computation in a short amount of time.

CLUSTER HARDWARE

The components of the Helsinki Institute of Physics *testbed1* prototype cluster frontend and six compute nodes were:

- CPU: Intel 866 MHz Pentium III Coppermine
- MB and case: Siemens Nixdorf
- Memory: 256 MB
- ATA disk: 10 GB Seagate ST310211A 5400 RPM 1 MB cache
- NIC: Intel 82801BA/BAM/CA/CAM
- CDROM and FDD: yes

- Switch: unmanaged 8 port Zyxel ES-108 100/10 Mb/s

The frontend has an additional 3Com PCI 3c905C Tornado NIC and a 15 GB IBM-DJNA-351520 5400 rpm 430 kB cache harddisk instead of a 10 GB disk. Worker node *compute-0-4* has a Seagate ST330610A 30 GB 5400 RPM 2 MB cache harddisk instead of a 10 GB disk. A snapshot from the monitoring tool shows the state of the *testbed1* cluster nodes in Figure 4.

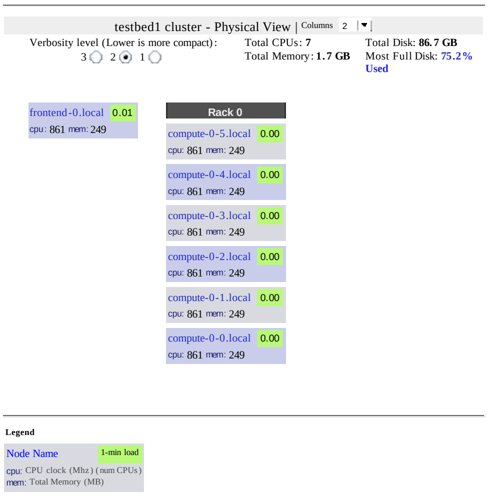

Figure 4: The CPU frequency, memory and load of the cluster nodes is shown with a web based monitoring tool.

PROOF

PROOF is an extension to ROOT which allows the parallel analysis of large ROOT trees [20]. The large data sets produced by present and future HEP experiments makes it a very interesting and useful tool. PROOF builds on the experiences gained by the Physics Analysis Workstation (PAW) team from the Parallel Interactive Analysis Facility (PIAF). PIAF run only on dedicated homogeneous farms, but PROOF can be set up on inhomogeneous desktop computers in a research group.

In this work the concept of a dedicated parallel interactive analysis cluster has been explored. Setting up a Linux cluster with standard inexpensive PC hardware is easier than before, using a Linux distribution specialized for cluster installation, configuration, monitoring and maintenance. For this, NPACI Rocks Cluster Distribution is one of the best. Cluster software should reduce the task of maintaining N+1 computers to maintaining only 1+1 computers. A good cluster distribution makes it trivial to add or remove nodes from the cluster. It also simplifies making software updates and changes on all nodes. This can make it easier to set up a dedicated PROOF cluster than using a lot of different workstations in a research group.

Setting up a prototype cluster

To try out PROOF we built *testbed1* using hardware recycled from an OpenMosix computer class cluster [21]. We chose Rocks as the cluster distribution because of good exeriences gained with it setting up a 64+2 CPU Linux cluster.

PROOF was installed on the frontend node and distributed with NFS to the compute nodes on the small Rocks test cluster. On a large cluster a local installation on each node could ease the performance problems NFS is know to suffer from.

We modify the Rocks XML configuration scripts to install PROOF also on the compute nodes. PROOF supports several secure and unsecure authentication methods to connect the worker nodes (UsrPwd, Secure Remote Password (srp), Kerberos (krb5), Globus X.509 certificates, ssh and uid-guid). Authentication of the worker nodes failed for unpriviledged users, until it was realized that the worker nodes need read and execution access to the NFS mounted home directories. Still there are remaining issues such as that ssh-authentication fails. At the moment documentation extending the basic README files is lacking and we are looking forward to the release of a PROOF manual.

CLUSTER PERFORMANCE

To get the maximum performance out of PROOF one should distribute the files to be analyzed over the local harddisks of the worker nodes. But PROOF can give a speedup even if the all files are stored on a central (NFS) server. This depends on the event sizes, the computational complexity of the analysis of each event, the CPU speeds and the I/O bandwidth of the used filesystem.

Figures 1-3 were computed with 1, 2, 4 and 7 worker nodes and the resulting event processing rates for Figure 1 are shown in the following table. In this case the performance increases up to four nodes, but does not scale linearly. This shows that PROOF works on our prototype cluster. The performance drop for seven nodes is probably because of the very limited I/O capacity of the single ATA disk on the frontend starts to decrease with increasing load.

Table 2: Exmple of PROOF performance when accessing data from a central NFS server instead from local worker node files.

Nodes	*Events/s*	Time [s]
1	1642	119
2	2471	80
4	3420	58
7	3088	64

CONCLUSION AND DISCUSSION ON VALIDATION PROJECT

We believe that the isotope production capability of the Geant4 Bertini cascade model can be used in many challenging applications such as HEP experiments at the Large Hadron Collider. To validate this, we have made extensive Geant4 simulations on proton-induced reactions with Pb and Au targets and performed PROOF based distributed data-analysis for these Monte Carlo events.

So far we have not found any serious software flaws in the Bertini cascade model. In comparison with experimental data general features are reasonably well reproduced, but many improvements can be thought of. Especially the use of a simple explosion model should be re-evaluated.

Following the example of extensive studies of the cascade-exiton model in [22], we plan to extend our Geant4 cascade analysis further towards more systematic treatment, and compare our results against other codes. Furthermore, we plan to study separately the sub-models and attack specific model problems. Also, further extensions and improvements of the Bertini cascade models are expected [7].

Rocks and PROOF are useful and powerful tools with a lot of potential, deserving more extensive documentation.

The experiences gained with integrating PROOF with Rocks on this small prototype cluster can be used to reach the full potential of parallel data analysis. Our present 64+2 CPU and future 132+2 CPU Rocks clusters will used for Compact Muon Solenoid production simulations and analysis. The idea of integrating PROOF with a batch queue system seems very interesting and could be the area of further work [23].

ACKNOWLEDGEMENTS

We highly appreciate the PROOF team who has been very helpful in debugging our configuration problems,

REFERENCES

[1] J. P. Wellisch A. Heikkinen, N. Stepanov. Bertini intra-nuclear cascade implementation in Geant4. In *Proceedings of CHEP03, La Jolla, California 24.-28.3. 2003. [http://arxiv.org/pdf/nucl-th/0306008. Also published in CD format SLAC-R-636*].

[2] J. P. Wellisch. Hadronic shower models in Geant4 – the frameworks. *Comp. Phys. Commun.*, 140:65–75, 2001.

[3] V.N. Ivanchenko. Geant4: physics potential for HEP instrumentation. *Nucl. Instr. and Meth.*, A(494):514–519, 2002.

[4] V. N. Ivanchenko. Geant4: physics potential for instrumentation in space and medicine. *Nucl. Instr. and Meth.*, (*Article in press*), 2004.

[5] A. Boudard et al. Improvements of intra-nuclear cascade model stimulated by recent spallation data. In *AIP Conference Proceedings*, volume 610(1), pages 300–304, 2002.

[6] D. H. Wright et al. Using Geant4 in the BaBar simulation. Technical report, CHEP03 proceedings, 2003. *[Available at http://www.slac.stanford.edu/pubs/slacpubs/9750/slac-pub-9862.pdf]*.

[7] D. Wright and A. Heikkinen. Adding kaons to the Bertini cascade model. In *Proceedings of CHEP04*.

[8] J. Beringer. (p, xn) production cross sections: A benchmark study for the validation of hadronic physics simulation at LHC. Technical report, CERN-LCGAPP-2003-18, 2003.

[9] Y. E. Titarenko et al. Study of residual product nuclide yields from 0.1, 0.2, 0.8, and 2.6 GeV proton-irradiated ^{nat}Hg targets. *LANL Report LA-UR-00-3600*, 2000. *[Available at http://arxiv.org/pdf/nucl-ex/0008012]*.

[10] V. N. Ivanchenko. The Geant4 hadronic verification suite for the cascade energy range. Technical report, CHEP03 proceedings, 2003. *[available at http://arxiv.org/pdf/nucl-th/0306016]*.

[11] J. P. Wellisch. Geant4 hadronic physics status and validation for large HEP detectors. Technical report, CHEP03 proceedings, 2003. *[Available at http://arxiv.org/pdf/nucl-th/0306006]*.

[12] J. P. Wellisch. Geant4 physics validation for large HEP detectors. *Nucl. Instr. and Meth.*, A(502):669–672, 2003.

[13] U. Georg et al. Isotope production comparison at ISOLDE with 1 and 1.4 GeV protons. *Nucl. Phys.*, A(701):137c–143c, 2002.

[14] K. Niita et al. Analysis of the (n, xn') reactions by quantum molecular dynamics plus statistical decay model. *Phys. Rev. C*, 52(5):2620–2635, 1995.

[15] K. Niita et al. High-energy particle transport code NMTC/JAM. *Nucl. Instr. and Meth.*, B(184):406–420, 2001.

[16] M. Enke et al. Evolution of a spallation reaction: experiment and Monte Carlo simulation. *Nucl. Phys.*, A(657):317–339, 1999.

[17] A. Boudard et al. A new model for production of fast light clusters in spallation reactions. *Nucl. Phys*, A(740):195–210, 2004.

[18] A. Letourneau et al. Composite-particle emission in the reaction p+Au at 2.5 GeV. *Nucl. Phys*, A(712):133–166, 2002.

[19] M. Ballintijn et al. The PROOF distributed parallel analysis framework based on ROOT. Technical report, CHEP03 proceedings, 2003. *[Available at http://arxiv.org/pdf/physics/0306110]*.

[20] R. Brun et al. ROOT status and future developments. Technical report, CHEP03 proceedings, 2003. *[arXiv: cs.SE/0306078 Available at http://arxiv.org/pdf/cs.SE/0306078]*.

[21] A. Heikkinen et al. The performance of a new Geant4 Bertini intra-nuclear cascade model in a high throughput computing (HTC) cluster architecture. In *Proceedings of SNA, Paris 22.-24.9 2003*.

[22] S. G. Mashnik et al. Cascade-exiton model analysis of proton induced reactions from 10 MeV to 5 GeV. *Nucl. Instr. and Meth.*, A(414):68–72, 1998.

[23] K. Schwartz et al. PROOF – Parallel ROOT facility – A use case. *[Available at http://www.slac.stanford.edu/BFROOT/www/Computing/-Distributed/ROOT2004/files/schwarz.ppt]*.