Development, validation and maintenance of Monte Carlo generators and generator services in the LHC era

P. Bartalini, CERN, Geneva, Switzerland
M. Kirsanov*, INR, Moscow, Russia

Abstract

The library of Monte Carlo generator tools maintained by LCG (GENSER) guarantees the centralized software and physics support for the simulation of fundamental interactions, and is currently widely adopted by the LHC collaborations.

While the activity in the LCG Phase I was mostly concentrating in the standardization, integration and maintenance of the existing Monte Carlo packages, more emphasis is currently being placed on the contribution to the development, testing and primary validation of new Monte Carlo packages, most of them designed using object oriented technologies.

The current status and the future plans of this activity are presented, with emphasis on the testing and validation.

The status of MCDB, the public database for the configuration, book-keeping and storage of the generator level event files is also reviewed.

INTRODUCTION

The project LCG Generator was started in 2003 as a long term project. Its purpose is to provide support to the high-energy event-generation software and related services in the LHC era. The project consists of four subprojects:

- WP1: GENerator SERvices library (GENSER)
- WP2: Event formats and event interfaces
- WP3: Shared event files: framework and Data Base (MCDB)
- WP4: Tuning and validation

The following scientific centers and groups participate in the project: Florida University, CERN, LCG Russia, LCG Spain.

The LCG generator information can be found at http://lcgapp.cern.ch/project/simu/generator/

GENSER

The purpose of this subproject is to replace the obsolete CERN library for what concerns the MC event generators and related software and to maintain them during a long period of the LHC operation, in conditions of appearance of new computer platforms and compilers. It should be noted that some of the packages that are included in GENSER are written in Fortran during a long period of time and can make use of old Fortran features. At the same time some of them are sufficiently complicated, and the corresponding Object-Oriented technology substitutions should be developped and validated for years before they reach production quality.

The duties of the GENSER team are the following:

- To collaborate with the MC authors in order to prepare LCG compliant codes
- To maintain older packages on the LCG supported platforms
- To perform the integration of new packages on requests from LHC experiments and other GENSER users.
- To help users in their work with MC generators

In 1004 - 2005 GENSER reached the production quality. ATLAS and LHCb MC productions rely on GENSER. At present 25 MC generators and related packages are included in GENSER. This number grows, but since 2005 not so quickly, there is some suturation. The accent in this subproject is being shifted towards **convenience, tests, validation**.

LCG GENERATOR TESTING AND VALIDATION: GENERAL CONSIDERATIONS

Large scale MC productions in LHC experiments involve a lot of manpower and computer resources. They are rather expensive. At the same time, a bug in a MC event generator, depending on its severity, can make such a production mostly useless. In order to avoid such problems, the MC generators must be thoroughly tested. There are, however, contradicting requirements to this validation: on the one hand, it should be thorough and detailed, but on the other hand, it should not bee too slow.

In the opinion of the LCG Generator team, several **levels** of testing and validation can help. These levels are seen as follows:

- **Level0**: Special script checks the existence of non-zero size libraries according to definite rules (one, several or no libraries should correspond to a package, the latter in case of external package) **Level1**: Dedicated package in GENSER. The test script compiles one or several applications for each generator to be tested and runs them one by one. Each application writes some numbers in a file. At the end the summary file is compared with the standard file and the differences are re-

* Mikhail.Kirsanov@cern.ch

ported **Level2**: Dedicated projects (like WP4) **Level3**: Tests inside the experiments

Note that levens 2 and 3 are outside WP1 GENSER. Additional levels can also be used.

LCG GENERATOR TESTING AND VALIDATION: LEVEL 1 TESTING

For this purpose the dedicated package TESTS in GENSER was created. The libraries of this package contain service routines and codes that perform some simple analysis and reconstruction. Under the latter we mean jet reconstruction and lepton isolation. The directory examples of this package contain the test scripts and main modules for the test applications. Some additional dependencies of GENSER like the dependency on ROOT are moved here (for the moment all other GENSER libraries do not depend on ROOT, but this will not be necessarily the case in future).

The tests procedure is defined by the fact that the results of the simulation are often **correct, but not identical** since we deal with **Monte Carlo** programs! For this reason all results are compared with the standard ones taking into account the statistical errors, that are always written in the output files together with the results.

A special program performs the comparison. All significant changes are reported as warnings (for the deviation above the statistical error value multiplied by 3) or errors (for the deviation above the statistical error value multiplied by 4). Missing results (this can happen if the script failed to compile an application or if an application crashed) are also reported without stopping the comparison program.

7 most important generators out of 25+ are tested at present.

The full testing takes about 1 hour. This takes into account that shared and archive libraries are tested separately, generators with HEPEVT common block with sizes 4000 and 10000 also separately.

In 2005 at least two bugs and unexpected features are found with level 1 GENSER testing, at least one bug in PYTHIA found with level 3 CMS. The plan is to take part of CMS testing in the GENSER package TESTS.

PREPARATION OF THE LCG GENERATOR TO NEW COMPILERS

The future version of the GNU compiler is gcc4. In this generation gfortran (essentially Fortran 95) substitutes g77. gfortran does not support all Fortran 77 features.

Although the platforms with gcc4 are not yet in the list of platforms supported by LCG, the GENSER team performed some work in preparation for this new generation of compilers. Small problems were encountered in PYTHIA, much more problems in HERWIG (return to label, entry points). In collaboration with authors the versions of these most popular general purpose Fortran generators were prepared and included in GENSER since the GENSER release 1_2_1.

64 bits platforms also are coming. Their introduction will require a thorough testing. Strange behavior of old Fortran codes on these platforms was already noticed.

LCG GENERATOR RELEASES AND PLANS

The last GENSER major release is 1_2_0.

After that a bugfix "light" release 1_2_1 was also made. In the "light" release the unchanged packages are symlinked to the previous release. The tools for the "light" release were developped by the librarian Andreas Pfeiffer. They are to be further developped.

The other main points of the plans of GENSER for 2006 are:

- Development of the package TESTS. The goal for 2006 is to increase the number of tests by at least a factor of 3 - 4
- Integration and further support of c++ generators Pythia8 (Alpha version available) and Herwig++ (Beta version)
- User support for releases 1_2_1 - 1_4_0
- New packages: DPMJET(?), PYQUEN, HYDJET
- Continuation of testing with gcc4
- Connection with MCDB

WP2: EVENT FORMATS AND EVENT INTERFACES

The goal of this subproject is to standardize interfaces, avoiding the proliferation of versions and duplication of work. This should improve conditions for the development and validation of the new MC generators created using Object Oriented technology.

One of the most popular interfaces is the MC truth interface HepMC. In January 2006 the baseline version was chosen and installed in the LCG external area. It is based on the version 1.26 from M. Dobbs. Interfaces to MC generators (the most unstable part of the package) are split into a separate library. Further development started.

The GENSER team plans to increase the number of tests that use HepMC interface. This means that new verions of HepMC will be tested during the GENSER level 1 tests.

The HEPML - Meta-data format facilitating automated documentation - is being prepared by the MCDB team. The parallel work is being performed under CEDAR. In the nearest future the common approach should be reached.

WP3: PRODUCTION AND MAINTENANCE OF SHARED EVENT FILES

The goal of this subproject is to produce, store and maintain the certified generator level event files. This subproject

is motivated by the increasing complexity of the event generation due to the increasing number of physical models, parameters of the models and future increase of accuracy requirements to the generated samples as the time of LHC operation approaches.

The corresponding production framework is being prepared in collaboration with CMS. The prototype is available.

For the storage and maintenance of these event files the Monte Carlo Data Base (MCDB) is developped under the subproject WP3. The maintenace includes book-keeping and providing documentation for the files, using the competences of the Monte Carlo experts. MCDB is currently in production, only a few thing are still to be finalized.

The next challenging task in this subproject is to populate MCDB and attract users.

WP4: TUNING AND VALIDATION OF THE MC GENERATORS

This subproject deals with specialized tools for the detailed testing, comparison with data and tuning of the generators. The following existing tools can be mentioned here.

- JetWeb http://jetweb.hep.ucl.ac.uk
- HzTool under CEDAR

The first one plans to switch to GENSER for what conserns the generators libraries. For second one the proposal to include it in GENSER is being discussed.

From the tools under development one can mention the generator analysis framework, prtly coinsiding with the production framework mentioned above.

Some standalone studies are also known, for example the ALPGEN validation in Perugia and HIJING validation at JINR.

CONCLUSION

In order to provide the MC generator services in the LHC era, the LCG generator project was started in 2003. The subprojects GENSER and MCDB of this project are now in production, the former one is already widely used by the LHC experiments. To provide the levels 0 and 1 testing of the generator library GENSER the special GENSER package is created. Its development continues. In other subprojects of the LCG generator devoted to the MC tuning and validation the development is underway.