Review of the Geant4 Project
April 16 – 20, 2007

Report submitted on April 30, 2007

Members of the Review Team:

Irène Buvat, INSERM
Fabio Cossutti, INFN Trieste
Frank Gaede, DESY
Nobu Katayama, KEK
Andreas Morsch, CERN
Robert Reed, Vanderbilt University
Sayed Rokni, SLAC
Lembit Sihver, Chalmers University of Technology
Charles Young, SLAC (chair)

1. Executive Summary

The committee is impressed by the progress by the Geant4 collaboration over the past several years. It has grown significantly beyond high energy physics (HEP), and is now a widely acknowledged simulation tool of choice in diverse application domains, including medical and space applications.

Existing physics models have been improved, and new ones have been added. The user can choose to use those models best suited to his specific application. Computing performance has kept pace despite the addition of these new features. Our major findings and recommendations deal with this impressive array of capabilities.

- There is widespread interest in the user community to have the FLUKA hadronics physics model available within Geant4.

- The two electromagnetic models should be treated in a single package, similar to hadronic models, to improve their physics utility and reduce user confusion.

- With the plethora of physics models, clear documentation of validity ranges and expert guidance are needed to help users make the right choices.

- The needs of the space and medical user communities often overlap with the HEP community. However, these communities have physics related needs that are different than those traditionally associated with the HEP user community. Enhancing this support, specifically improvements in the hadronic physics area, will help the continued strong growth of Geant4 in these areas.

- Computing performance should be improved so that users can take advantage of the flexibility of Geant4 simulations for practical applications in which computing time is a limiting factor, such as HEP, medical and space applications.

- A manpower and resource plan is needed in order to achieve these improvements in a timely manner. Milestones should be driven by the “need by” dates from the user community.

2. Introduction

This report contains the result of the review of the Geant4 project held at CERN on April 16-20, 2007. The Geant4 Oversight Board appointed members of the review committee and provided its mandate, which can be found in Appendix A.

The committee solicited input from the user community over a period of more than a month. A summary was provided to the Geant4 collaboration prior to the review.

Petteri Nieminen, chair of the Geant4 Oversight Board, reviewed the charge at the beginning of the review. John Apostolakis, spokesperson of the Geant4 collaboration, began the review with a brief summary of collaboration activities. This was followed by six reports summarizing user inputs. There were two and a half days of presentations from the Geant4 collaboration addressing these and other issues. A summary of findings and recommendations was presented on the last day. The complete agenda is included in this report as Appendix B.

The committee would like to thank the Geant4 collaboration and CERN for their organization and support of the meeting, and for their hospitality.

The Geant4 project has been a tremendous success since its inception. The Geant4 toolkit quickly gained acceptance as the tool of choice by high energy physics (HEP) experiments at the Large Hadron Collider (LHC) and a number of HEP experiments elsewhere. It has now become a tool of choice in the medical community and the space science community as well, and it has been considered for use in other disciplines such as neutron instrumentation. The physics models within Geant4 have grown to cover a larger range of applicability. The capabilities of Geant4 are particularly well suited for the complex geometries in these applications. The review committee commends the entire collaboration.

3. Physics Performance

The scientific goal of a simulation program is to reproduce with great accuracy what can be measured. There are several factors that can contribute to loss of fidelity.

- Inadequate physics model for the underlying process. It could be the interaction of different types of particle with detector material in an HEP experiment. It could be the interaction of cosmic ray particles with space craft electronics or human organs. It could be the interaction of photons, particles or heavy ions with human cells during radiation therapy. Physics models are provided in the Geant4 toolkit.

- Inaccurate description of target material. It could be the chemical composition of a HEP detector element. It could be the thickness of metallization in electronics. It could be the size and shape of cells. Users are responsible for using the tools provided by Geant4 to describe the target. It is then the toolkit's job to utilize this geometry properly, e.g. whether an X-ray is within a cell or not.

- Incorrect detector response model. The immediate physical outcome of the interaction is often experimentally inaccessible, and the physical observable is a derived quantity. For example, there may be electronic amplification of a signal with the introduction of some noise. This part is entirely the user's responsibility.

Physics performance refers to the first factor. On the other hand, any discrepancy between simulation and data could be due to any one or more of these factors. It is therefore important to have good control over the other two factors in order to draw conclusions on physics performance.

3.1. Electromagnetic Physics

There are two electromagnetic (EM) packages in Geant4: standard EM and low energy EM.

The standard EM group reviewed its recent developments. These include an improved treatment of multiple scattering, single elastic scattering, ionization processes for hadrons and ions, improvements in Compton scattering, synchrotron radiation and transition radiation. There is a new library for polarized processes. Extensive comparisons show that the model agrees well with data over large kinematic regions, from 100's keV to 100's GeV in energy for some processes. The improved multiple scattering model has resulted in a significant slowdown of the code. The group has provided an option for cases where the improved physics precision cannot justify the additional CPU time.

The low energy EM group's presentation focused on its validation strategy. The model covers electrons and photons in the energy range ~100 eV to ~100 GeV; it contains shell effects, and it has atomic relaxation processes such as fluorescence and Auger emission. The low energy EM package is to be extended to the eV scale micro-dosimetry applications.

Users have reported that documentation, such as model validation and comparison with data, can be hard to find. Newer heavy ion stopping power models are not fully available. Users appreciate the ability to make tradeoffs between physics accuracy and computing speed; however, they are frustrated by the difficulty of finding the optimal point in a large parameter space. The committee recognizes that there is no one-size-fits-all choice of parameter values that would apply to all cases.

Recommendation 1: We recommend setting up an easily accessible central repository with detailed references to data comparisons and validation papers for the models, e.g. a web page linked from the EM home page.

Recommendation 2: We recommend rapid integration of the ICRU 73 heavy ion stopping power model.

Recommendation 3: We recommend providing guidance on the tradeoff between physics accuracy and computing speed. For example, collect and publish via the web results from users in various domains regarding computing speed vs range cut, and provide a Wiki-like forum where users can document their experience; etc.

Users have been confused by the existence of two independent EM packages. They have overlapping ranges, where they give different results for identical geometry. Which package gives the more correct result in this kinematic range?

Recommendation 4: We recommend providing guidance on the choice between the two EM models for specific particle species, energies, etc.

The two models do not have identical ranges of applicability. It is therefore natural to want to use one model in one region and the other model in another region, each in a range where it has better performance. This is impossible as the models have been implemented as independent packages.

Recommendation 5: We recommend integrating the two EM models into a single package, similar to what exists in hadronic models. This will allow a user to choose one model in one energy range and the other model in a different energy range in order to optimize physics and computing performance for his application.

3.2. Hadronic Physics and FLUKA

The committee has received strong requests from many segments of the user community to be able to use the FLUKA hadronic physics models within Geant4. Similar comments were expressed at previous Geant4 reviews.

The Geant4 toolkit and the FLUKA package have different strengths. Geant4 is much more capable in describing the complex geometries common in HEP, space and medical applications. It has a standard user interface for hit generation. And it is an open source toolkit. On the other hand, many users have identified FLUKA as best suited for their applications due to its mature and carefully benchmarked hadronic physics model. Some HEP experiments that would like to use Geant4 have not done so because of the absence of FLUKA physics models in Geant4. The simulations involved in radiation shielding of accelerators have long used FLUKA. Their interest in Geant4 has grown as their geometries get more complex, but they are reluctant to switch to Geant4 without access to the FLUKA physics model.

The FLUKA Virtual Monte Carlo (VMC) implementation developed by the ALICE collaboration is a possible solution to this problem. There are two drawbacks.

- The natural time to choose a framework is at the beginning of a project. Many HEP experiments already have Geant4-based simulation frameworks. Switching to VMC at this time would be highly disruptive.

- While the VMC approach allows one to use FLUKA physics models with Geant4 geometry, it does not have the physics flexibility that many users require. In particular, it does not allow the possibility of combining FLUKA physics models with Geant4 physics models in one application, for example to get the best performance for various energy regimes.

The committee understands that there have been contacts between the two groups. The committee also recognizes the difficulties of integration, both technical and otherwise. Nevertheless, the committee feels that it is crucial to continue to pursue a solution as Geant4 and FLUKA both have an interest in satisfying this strong and continued user request.

Recommendation 6: We recommend continuing the dialog with the FLUKA collaboration to find a cooperative way to provide users with the best features of Geant4 and FLUKA. One possible implementation could be a loosely coupled interface between Geant4 and FLUKA hadronic physics model.

3.3. Other Hadronic Physics Issues

Hadronic shower simulation is important to many applications. For example, LHC experiments demand a good understanding of the longitudinal and radial shower profiles in order to properly simulate their detectors and hence physics capabilities. Detector development for the International Linear Collider (ILC) envisions a calorimeter with very high granularity optimized for particle flow measurements in a multi-jet environment. It can potentially distinguish the sub-showers within a single hadronic shower.

Geant4 has made major strides in hadronic simulation in the past few years. There are many physics model: low and high energy parameterized model (LHEP), quark gluon string model, Fritiof model, binary cascade, Bertini cascade, chiral invariant phase space model (CHIPS), and so on. Many validation results were presented for individual models.

Geant4 allows the use of multiple models in an application. This is achieved through the use of a Physics List (PL), where one specifies the models to use and their applicable particle types and energy ranges. Thus, the user can customize and optimize for his specific application. The committee applauds this flexibility.

However, this flexibility also presents a problem: the user is faced with a bewildering array of possibilities. Geant4 has addressed this with a number of suggested PLs that are likely to be reasonable choices in various domains. The committee commends Geant4 for providing this guidance.

There has been a large effort on validation since the last review. This has been undertaken by the Geant4 collaboration itself and together with others such as the LCG Physics Validation group and LHC experiments. These validation tests can be divided into two categories.

- Thin target. These tests are geared towards testing specific physics models and processes, and often in limited kinematic ranges.

- Thick target or full setup. These test PL in the context of real-life experimental setups, e.g. the test beam calorimeters of the LHC experiments ATLAS and CMS.

Some important shortcomings of the Geant4 hadronic models have been uncovered. The simulated hadronic showers start and stop too early compared with LHC calorimeter test beam data, and they are too narrow in the transverse direction. Geant 4 is aware of problems in the hadronic models in the few GeV/c range, which in turn might give rise to the observed disagreement of shower shapes at higher energies. Possible solutions were presented during the review. However, there were no detailed plans with milestones for the improvement of the hadronic package.

Recommendation 7: We recommend that Geant4 develop detailed plans to improve the hadronic package, in particular to improve the simulated shower shapes.

The committee notes that it might have been possible to uncover these issues and start addressing their underlying causes much earlier by comparing with well known shower shape parameterizations.

Recommendation 8: We recommend putting in place a set of simple hadronic benchmarks which allow quickly identifying very basic problems like disagreement with well known shower shapes.

In the space community, there is a strong need to have an accurate prediction of the triple differential cross sections ($d^3\sigma/(dEdZd\theta)$) for recoiling particles from all interactions, including heavy-mass ion-ion reactions.

When using heavy ions for radiation therapy – currently carbon is the dominant species but the use of other ions is under active investigation – there is also a need for highly accurate prediction of triple differential cross sections.

Recommendation 9: *We recommend that Geant4 accelerate the development of physics models to address these needs.*

4. Computing Performance

Computing performance affects all users, and is a particularly serious concern for heavy users such as HEP, medical and space applications. Improvements in computing performance translate directly to the amount of hardware that is needed, and therefore translates to real money and effort.

It should be noted that computing performance is more than just the execution speed of a piece of code. It is the scientific impact the end user experiences for a given investment in computing resource. A multi-prong approach is needed.

- Code optimization. Geant4 has profiled its code, and "hot spots" have been addressed. The committee is impressed with this tremendous effort. While profiling is a powerful tool able to identify localized problems, it cannot address all performance issues. Structural issues may require scrutiny by computing experts; physics related issues may require physicist involvement. For example, a particular physics model has seen a factor of two improvement in speed as a result of computing professionals going over the code.

  Recommendation 10: We recommend extending the scope of the computing professionals to review and optimize all Geant4 code.

  In general, the profiled applications do not have the same coverage and usage as real-life applications, i.e. "hot spots" can appear in user applications that are not apparent during testing.

  Recommendation 11: We recommend that Geant4 encourage users to monitor their applications, and provide feedback so additional "hot spots" can be identified.

- Trade-off between simulation detail and CPU time. In the case of applications that do not need the ultimate achievable physics accuracy, one can opt for a cruder but still acceptable level of physics accuracy and gain on CPU time. This is a choice for the user; however, Geant4 can play a crucial role to help the user choose the best physics model(s), parameters values as well as geometry description, and hit scoring methods.

  Recommendation 12: We recommend the creation of a performance optimization guide. It is likely that such information already exists and just needs to be collected into one document.

  A special case of this kind of trade-off involves turning off processes that are "irrelevant" for a given region, while retaining the other physics features of the model. There are use cases where this can save significant CPU time with little or no physics impact.

Recommendation 13: We recommend providing a simple mechanism for users to turn off "irrelevant" processes for a given region.

- Variance reduction technique (VRT) or event biasing. VRT can greatly enhance the impact of CPU power in cases where certain low-probability outcomes are important to the application. Geant4 provides tools for VRT; however, the user community appears to be not well informed of these tools or how to use them.

  Recommendation 14: We recommend providing guidance on the use of VRT in Geant4 so users can better take advantage of them. References to publications and data regarding the validation of the VRT methods in Geant4 should be made available in the performance optimization guide.

  Reverse Monte Carlo is a variant of VRT, and can be important in some applications. Geant4 should clarify its plans on this.

Computing performance is not a one-time issue, but requires on-going attention. Performance will change as the code is updated. Performance is affected by new hardware architectures and compilers. Overall performance changes as new physics models are introduced.

Recommendation 15: We recommend systematic tracking of code performance for each part of the code, and for each physics model. Comparisons with previous versions should be an integral part of the release notes.

Many computers now come with 2- or 4-core CPU. Industry projection is for having many more cores, perhaps hundreds per CPU. Running multiple threads per application can speed up the turn-around of any single job, and the same overall throughput can be achieved with less memory than utilizing one core per application.

Recommendation 16: We recommend that Geant4 keep itself abreast of developments in the area of multi cores and advanced instructions, so it can take advantage of them when there is sufficient infrastructure and support to do so.

While the unit cost for computing has gone down, computing requirements have gone up, so resource needs continue to rise. Projects often have to make plans several years in advance. It is important to give them the best information for such projection.

Recommendation 17: We recommend that Geant4 publish a plan regarding the expected computing performance of the toolkit over the next five years.

5. Usability Issues

Usability has many components. They are all the things that allow a user to develop a science application from the Geant4 toolkit and then apply it to their problems, with great confidence in the results and without large overheads. These include documentation, e.g. installation and setup procedures and physics validation, visualization tools, tutorials, user workshops, and so on.

5.1. Physics Validation

Geant4 provides a range of physics models, which can be combined to form the PL for a particular application. There is an extensive program to validate the models. Geant4 has started a database of experimental data for this purpose; however, the comparisons to date between data and models have been limited. There is also not a systematic program of comparison with other Monte Carlo transport codes.

Recommendation 18: We recommend aggressively populating the database with all relevant experimental data, as well as validation results provided by others.

Recommendation 19: We recommend continuously and systematically benchmarking against other Monte Carlo transport codes.

5.2. Release Validation

There is one major release and one or two minor releases each year. Software quality assurance (QA) checks cover the many computer platforms and compilers that are supported. In addition, users are invited to participate in beta testing. This testing process takes approximately 3 weeks. The results are examined manually and compared with previous releases. This comparison is therefore at times qualitative and subjective. Physics performance and computing performance are tracked, but in an ad-hoc fashion.

It is not possible to predict how some future problem will manifest itself. It is therefore important to monitor a comprehensive set of variables. The committee believes the comparisons should be automated and made quantitative. Automation will insure consistency from release to release, allow the tracking of many more parameters, and allows the Geant4 physicists to concentrate on the evolution of the physics results with release rather than the much less interesting task of producing the validation results.

Recommendation 20: We recommend defining and automating a common validation procedure to be run for every release, monitoring a comprehensive set of variables and exploiting the comparisons with the collected experimental results.

It is important to quantify and then track the validation results. The committee recognizes that some properties may not have an unambiguous way to quantify. On the other hand, one can provide multiple quantitative measures for a property, e.g. one could characterize the transverse dimension of a shower by the Gaussian σ of its core, by the RMS of the entire distribution, by the fraction of data not represented in the Gaussian, and so on. Such a procedure does not in any way preclude archiving the distributions for further manual scrutiny as necessary

Recommendation 21: We recommend defining quantitative metrics for validation results.

Recommendation 22: We recommend that all validation results, both the quantitative metrics and the underlying distributions, be made easily accessible to the user.

5.3. Documentation

Geant4 currently provides extensive documentation, e.g.

- Application Developers Guide
- Installation Guide
- Getting Started
- User's Guide for Toolkit Developers
- LXR reference guide
- FAQ

A new effort is underway to move into the DocBook format for most documents, and to use MathXML for the physics manual because of its extensive use of equations. The committee commends Geant4's documentation effort, and encourages the expeditious migration to DocBook and MathXML.

Maintaining all this documentation is a large and continuing job. It is not surprising that a number of documents are out of date, and some web links are broken.

Recommendation 23: We recommend updating the documentation on the web right away, and formulating a plan to have periodic reviews of the web site to keep it up to date.

Users have reported the lack of clear high level documentation on physics models. For example, which of the two EM models should be used under what circumstances, what is the validity range of a PL, and so on. Users have also requested information on the connection between the physics models and their implementation.

Recommendation 24: We recommend that Geant4 document the limitations, and validity and applicability ranges of the different EM and hadronic *models, and the physics lists. Where models have overlapping validity ranges, document the tradeoffs.*

Code documentation appears to be not as well developed as in some other areas, and should be improved.

Recommendation 25: We recommend that Geant4 improve its code documentation. Example approaches by other groups include Doxygen, and README files that can be browsed in LXR.

The many documents have overlapping topics, and users are uncertain which one to consult. There is already a plan to combine some of these documents, and reduce the total number. The committee agrees with this, and looks forward to the early completion of this project.

Release notes are an important tool for users working on the state of the art version of the code. Currently, their contents are focused on the code, with limited explanation of the physics and performance issues which are often of greater interest to the user.

Recommendation 26: We recommend improving the release notes with the addition of expected changes in physics and computing performances, while keeping the current extensive list of code change descriptions. We encourage strengthening the editorial coordination of the release note preparation.

Some users have been intimidated by the installation procedure. The committee recognizes that it is extremely difficult to have a procedure that is on the one hand simple and is on the other hand applicable to all the situations experienced by the diverse user community. In fact, Geant4 has a fairly straight forward installation procedure. This is especially true for the larger user groups which have significant computing as well as Geant4 expertise. However, it is daunting for novice users to answer so many questions, even with the help of the installation guide.

Recommendation 27: We recommend that Geant4 review the current installation procedure with the aim of adding conveniences for users, such as additional defaults, self configuring procedure, a template `.spec` *file to create an RPM in the user's environment.*

5.4. User Support

Geant4 has organized many tutorials and workshops, some targeted at specific application domains. The committee has received extensive positive feedback. We applaud this tremendous effort and look forward to its continued success.

Users have expressed a desire to help. If documentation is editable like in a Wiki, then users can more easily contribute to it. This could be the fixing of minor errors. But perhaps more importantly, this could be the clarification of points that might be obvious to the expert. The committee recognizes the concerns with Wiki, such as the possibility of malicious editing, and well intentioned but incorrect information. However, many groups have successfully deployed Wiki for similar purposes without such problems. Some have simple access control to weed out the miscreant. The committee notes that correcting information on a Wiki makes the wrong information inaccessible. This is in contrast to HyperNews where the wrong information remains online and available to the unsuspecting, with the eventual corrections buried in long threads.

Users have found the interaction with one another during tutorials and workshops to be valuable. However, they also find this exchange often stops after the meeting, especially for people from different application areas. The existing HyperNews forums can be used for this purpose. However, the popularity of Wiki in the wider community makes it a more familiar interface.

Recommendation 28: We recommend setting up a user community supported Wiki as a collaborative tool. Geant4 should advise which documentations are appropriate to be moved to it.

While Geant4 provides an extensive set of physics models, it would be unreasonable to expect it to provide models in every kinematic regime and application domain. We commend Geant4 for the support it has provided to implement and integrate user driven physics models. It is inevitable that long-term support and maintenance will fall on the Geant4 collaboration. This is as it should be. Thus, the discussion of any particular new model should be driven by scientific needs only.

Recommendation 29: We recommend that Geant4 continue its vigorous support to implement and integrate user driven physics models.

6. Other Issues

The Geant4 collaboration has developed a powerful toolkit for simulation. It has been adopted by many application domains as the tool of choice: HEP, medical and space among others. And within each domain, the range of applications is also growing, e.g. the DNA project in the medical domain. The breadth of these programs and the mission critical nature of their applications have led to a large burden on the Geant4 collaboration. Its many tasks include the maintenance and continued enhancement of the core software, new and improved physics models, user support and so on. Manpower is a concern.

It is reasonable to expect the user community to provide some of this manpower. The development of physics models is a good match to user interest and abilities; however, the maintenance of these models is likely to remain with Geant4. Users can support one another, particularly on the more common usage issues, but Geant4 needs to furnish the infrastructure to enable this user dialog. There will be tasks that will remain as direct Geant4 responsibility, such as the core software and the basic documentation. The committee is therefore concerned that the available manpower may be insufficient to handle all these tasks in a timely manner.

Recommendation 30: We recommend that Geant4 develop manpower and resource plans for the next five years. They should be consistent with timely delivery of features and capabilities to the user community.

Recommendation 31: We recommend that Geant4 seek the additional support implied by the manpower and resource plans from all available sources.

7. Conclusions

The Geant4 project has been a tremendous success. Its usage has expanded beyond the traditional high energy physics domain to include medical and space applications. The collaboration has fostered an enthusiastic user community who are eager to participate in further development and support.

The future success of Geant4 depends crucially on the collaboration's ability to respond to the needs of this user community: physics capability, computing performance, and general usability. To this end, Geant4 should strive for the greatest possible openness and transparency to users, and the best possible customer support.

Appendix A: Charge

The mandate of the Geant4 Review Committee in 2007 is to investigate the physics precision, computational speed and general usability of the Geant4 software for the major existing and upcoming Geant4 use cases in its various application domains, and to issue recommendations for future software improvements and developments. In performing this task, the Committee may also address auxiliary issues that it considers relevant.

Appendix B: Review Agenda

All the presentations are linked to the review web page at
http://indico.cern.ch/conferenceDisplay.py?confId=14946.

April 16, 2007

14:15	Introduction and mandate	Petteri Nieminen
14:45	Geant4 Overview	John Apostolakis
15:15	User presentation (LHC physics)	Tancredi Carli
16:15	break	
16:45	User presentation (shielding)	Sayed Rokni
17:30	Executive session	

April 17, 2007

09:00	Executive session	
09:30	User presentation (ILC & other HEP physics)	Erika Garutti
10:30	User presentation (HEP usability)	Gloria Corti
11:30	break	
12:00	User presentation (medical)	Nicolas Karakatsanis
13:00	Lunch	
14:00	User presentation (space applications)	Robert Weller
15:00	break	
15:30	Standard EM physics overview	Vladimir Ivanchantko
16:10	The Fano cavity precision test	Michel Maire
16:50	Low energy EM physics	Maria Grazia Pia
17:30	Executive session	

April 18, 2007

09:00	Executive session	
09:30	Hadronics overview	Aatos Heikkinen
10:15	Hadronic validation Part 1: Thin target	Dennis Wright
11:00	break	
11:30	Shower shapes	Alberto Ribon
12:15	Overview of developments: CHIPS	Mikhail Kosov
13:00	Lunch	
14:00	Physics lists and hadronic physics	Gunther Folger
14:45	Hadronic validation Part 2	Dennis Wright
15:30	Hadronics: other issues	Alex Howard
16:00	break	
16:30	Executive session	

April 19, 2007

09:00	Executive session	
09:30	Geometry: performance, usability, tools	Gabriele Cosmo
10:00	Performance for space applications	Giovanni Santin
10:15	CPU performance – EM	Vladimir Ivanchantko
10:30	CPU performance – benchmarks	Alberto Ribon
11:00	break	
11:30	Release notes, beta testing and QA, External interface	Gabriele Cosmo
12:00	User-submitted classes/modules, documentation	John Apostolakis
12:15	Event biasing and scoring	Makoto Asai
12:30	Usability: visualization and UI	Joseph Perl
13:15	Lunch	
14:15	Executive session	
15:45	break	
16:00	Follow-up on physics	
17:30	Follow-up on usability	
18:00	Executive session	

April 20, 2007

09:00	Executive session
11:30	Private close-out with Geant4 management
12:00	General close-out
13:00	Adjourn