GEANT-4 GPU Port:

User Manual

Stuart Douglas – dougls2
Matthew Pagnan – pagnanmm
Rob Gorrie – gorrierw
Victor Reginato – reginavp

Version 0
February 28, 2016

Contents

1 Revision History

All major edits to this document will be recorded in the table below.

Table 1: Revision History

Description of Changes	Author	Date
Initial draft of document	Matt	2016-02-26

2 List of Figures

Table #	Title

3 Definitions and Acronyms

Table 2: Definitions and Acronyms

Term	Description
Geant-4	open-source software toolkit used to simulate the passage of particles through matter
Geant-GPU	Geant-4 with some computations running on the GPU
G4-STORK	(Geant-4 STOchastic Reactor Kinetics), fork of Geant-4 developed by McMaster's Engineering Physics department to simulate McMaster's nuclear reactor
GPU	graphics processing unit, well-suited to parallel computing tasks
CPU	computer processing unit, general computer processor well-suited to serial tasks
CUDA	parallel computing architecture for general purpose programming on GPU, developed by NVIDIA

4 Introduction

4.1 Purpose

4.2 Scope

4.3 Background

4.4 Document Overview

This document goes over how to install and run Geant4-GPU. As well as software and hardware required for Geant4-GPU to be installed and run. Following this are detailed step-by-step instructions on how to install the software. Following details on installation is a section on creating and running simulations using the CUDA functionality introduced in Geant4-GPU. A section detailing high level details about how the CUDA code was developed and integrated follows, guiding future developers and users on how to integrate CUDA into other parts of Geant4. A troubleshooting section is included as well as a FAQ section covering many common questions related to the project. Finally, the appendix includes conclusions about the state of the project and a brief description of the future of the project.

5 Legal Information

6 Installation

6.1 Recommended Knowledge

6.2 Required Hardware

6.3 Required Software

6.4 Supported Operating Systems

6.5 Installation Instructions

7 Execution

Geant4 is not an executable program in the traditional sense. It is instead a large set of libraries, designed to work together and that give you, the user, a framework to develop simulations for your work. The development of such simulations is not within the scope of this manual, but it is well-documented with a thorough guide available at `http://geant4.web.cern.ch/geant4/UserDocumentation/UsersGuides/ForApplicationDeveloper/html/index.html`.

This section will instead outline how one would go about using (or not using) CUDA computations with an existing simulation. The `Hadr04` example that comes with Geant4 will be used as the simulation to demonstrate this.

7.1 Enabling/Disabling CUDA

The installation guide details how to build the Geant4 libraries that will be used by your simulation. The Cmake command from step **??** includes a flag that determines whether or not CUDA will be used – `-DGEANT4_ENABLE_CUDA=ON`. Changing this value to `OFF` will cause Geant4 to build without any CUDA features, and without requiring `nvcc`, the CUDA compiler. After step **??** finishes, users will have to rebuild Geant4 by executing the `make install` command from their `geant4.10.02-build` directory. To re-enable CUDA, simply change the value to `ON` and rebuild again.

7.2 Building the Simulation

After Geant4 has been built, building the simulation is very straightforward. First, navigate to your simulation's root folder (`geant4.10.02/examples/hadronic/Hadr04` in our case) in your terminal. Then, create a new directory `bin`, so your Hadr04 folder now has two folders, `bin` and `src`. Go into the `bin` folder and run `cmake ../src` followed by `make`. That's it!

7.3 Running the Simulation

There are two main methods of running a Geant4 simulation. The first is to simply run the executable, in our case by running `./Hadr04` in the Hadr04 directory. This launches an interactive Geant4 command prompt, and you can manipulate all aspects of the simulation from within it, including the types of particles, their number, and their energies. This also allows you to visualize the simulation using the `vis` commands.

The alternative is to create a macro file that contains a sequence of commands that would be manually given to the interactive prompt, making it much easier to run a given simulation more than once and without requiring user interaction. Both methods print the results of the simulation out to the terminal window, as well as the running time of the simulation. If you wish to save the results to a file, that can easily be done via redirection in Unix. For example, if you wish to save the results of your simulation using `myMacro.mac` to `results.txt`, simply run `./Hadr04 myMacro.mac > results.txt`.

8 Porting Other Geant4 Modules to CUDA

There is currently one class in Geant4-GPU that uses CUDA – G4ParticleHPVector. This class was originally chosen as it was thought to be well-suited to parallelism. Although certain functions within the class are, the majority of the time spent within the class is not in these functions, the main reason for the performance degradations when CUDA is enabled.

There is indeed the possibility that there are other classes within Geant4 that are better suited to parallelism. This section will outline the general methodology we used to port G4ParticleHPVector with the hopes that it may be of use to future developers or users.

8.1 Layout of Source Files

An important consideration during development was to keep the CUDA code separate from the main Geant4 codebase as much as possible. We created a folder named `cuda` in `geant4.10.02/source/externals` to contain all CUDA source code. The `cuda` folder contains two folders, `include` and `src`. All header files are in `include`, and the CUDA source files are in `src`.

8.2 CMake Changes

Geant4 uses CMake as its build system, and minor modifications need to be made to integrate with CUDA. This consists of adding a variable at the top level CMakeLists file to enable or disable CUDA, and creating a Geant4_Add_Library_Cuda macro to Geant4MacroLibraryTargets. This uses `cuda_add_library` function included in Cmake. More details can be found in the project's detailed design document.

8.3 Interfacing with Existing Code

We used compiler directives within the default G4ParticleHPVector.cc file to make use of the CUDA code or to use the original implementation based on the flag passed to Geant4 during the build phase. If the flag is true, an object of type G4ParticleHPVector_CUDA is initialized and all function calls become one line calls to the corresponding function on the G4ParticleHPVector_CUDA object. For example, the `GetY` function looks like the following:

```
G4double GetY(int i) {
        #if GEANT4_ENABLE_CUDA
                return cudaVector->GetY();
        #else
                < existing code from G4ParticleHPVector.cc >
        #endif
```

```
}
```

8.4 Naming Conventions

It is important to use naming conventions to ensure clarity within a project that contains many possible ambiguities. The two main conventions we used were adding a "_CUDA" suffix to all source files within the `cuda` directory as well as to the class name of any classes implemented in CUDA. In addition to this, within all CUDA files there are two main types of pointer – those to GPU memory and those to CPU memory. To distinguish between these, all pointers to main memory are prepended with "h_" and pointers to GPU memory with a "d_".

9 Troubleshooting

9.1 Installation

- Ensure path names are correct in Cmake command
- Spaces in path names may cause issues on some systems
- The paths in step **??** of installing Geant-4 paths must be absolute
- Newer versions of Clang (included with Xcode 7) have been know to cause problems. Download Xcode 6 and uninstall Xcode 7 if this is the case.

9.2 Running the Simulation

- Ensure that your graphics card meets the requirements detailed in section **??**

9.3 FAQ

9.3.1 What is Geant-4?

Many physics researchers use Geant-4 to learn about how particles interact with a specific environment. It is a toolkit (i.e. set of libraries) that uses the Monte Carlo model, meaning each particle's properties are calculated independently according to certain probabilities. Users develop simulations based on these libraries.

9.3.2 Why will a GPU improve the performance?

GPU's contain a large amount of cores that can perform calculations much more quickly than a CPU if the problem is well-suited to parallelization. Geant-4 runs relatively simple calculations on millions of particles, and each particle is completely independent of the others. This is exactly that sort of well-suited problem, and stands to see large performance gains.

9.3.3 Where can I find more information about Geant-4?

For more information, checkout the official website at `http://www.geant4.web.cern.ch/geant4/index.shtml`

9.3.4 Helpful Pages:

- Download page for Geant4 source code: `http://www.geant4.web.cern.ch/geant4/support/download.shtml`
- Getting Started guide: `http://www.geant4.web.cern.ch/geant4/support/gettingstarted.shtml`
- Installation Guide for Geant4 (does not include CUDA): `http://www.geant4.web.cern.ch/geant4/UserDocumentation/UsersGuides/InstallationGuide/html/index.html`

10 Appendix

10.1 Recommendations for Integration of Geant4 and CUDA

10.2 Future of Geant4-GPU