

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 14日現在

機関番号：82118

研究種目：若手研究(B)

研究期間：2010～2012

課題番号 22740182

研究課題名（和文）GPUを用いた測定器シミュレーションの高速化技術の研究

研究課題名（英文）GPU parallel computing on detector simulation

## 研究代表者

村上 晃一（MURAKAMI KOICHI）

大学共同利用機関法人高エネルギー加速器研究機構・計算科学センター・研究機関講師

研究者番号：10353369

## 研究成果の概要（和文）：

本研究は、高エネルギー物理学実験で使われる測定器シミュレーションプログラムである Geant4 を、GPU コンピューティングの新しい手法を使って高速化することを目的とする。開発にあたっては、NVIDIA 社の Tesla C2070 / K20 ユニットを用い、GPU プログラミング環境である CUDA を使ってシミュレーションの並列処理化を行った。Geant4 の電磁相互作用を GPU 上で実装し、アプリケーションの速度向上を評価した。CPU と比較して、GPU で 30～40 倍程度の速度改善に成功し、研究計画段階で設定していた目標を実現した。

## 研究成果の概要（英文）：

Geant4 is a comprehensive particle physics simulation package that tracks particles passing through and interacting with matter. As more detailed and complex simulations are required in different application domains, there is much interest in adapting the code for parallel architectures. This work is a CUDA port of the core Geant4 algorithm, and tracks many particles in parallel on the GPU. The capabilities of GPU architectures provide an opportunity to enhance the performance of Geant4 applications. Benchmark tests shows 30-40 speedup compared to the single-threaded CPU version.

## 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	2,000,000	600,000	2,600,000
2011年度	500,000	150,000	650,000
2012年度	600,000	180,000	780,000
総計	3,100,000	930,000	4,030,000

## 研究分野：数物系科学

科研費の分科・細目：物理学、素粒子・原子核・宇宙線・宇宙物理

キーワード：素粒子実験、測定器シミュレーション、GPU コンピューティング

## 1. 研究開始当初の背景

Geant4 は、放射線と物質の相互作用をシミュレーションするためのソフトウェアである。高エネルギー・原子核実験、医療、宇宙といった幅広い分野で広く利用されてきており、測定器シミュレーションの重要なソフ

トウェアとなっている。近年、様々な応用分野において、精度の高いシミュレーション計算が要求されるのに伴い、シミュレーションの高速化は重要な課題となっている。

一方で、画像処理専用プロセッサである GPU (Graphics Processing Unit) を汎用的

な計算処理に利用する「GPU コンピューティング」が新しい計算機技術として注目されている。GPU は数百個の演算プロセッサを搭載し、CPU の 30~50 倍の処理性能と低コストを誇る。その演算性能は単体でもテラ FLOPS (1 秒間に行う浮動小数点演算の回数) のレベルに達している。スーパーコンピュータ並の計算を低コストでよりコンパクトなシステムで実現できる。ソフトウェアの開発環境も整備されてきつつあり、今後の新しい計算機技術として期待される。

GPU による Geant4 の測定器シミュレーションはアイデアとしてはあるが、実現には至っていない。GPU でのプログラミング手法は、CPU の逐次処理とは全く異なり、プログラミング上の制約も多い。GPU コンピューティング技術は発展途上の段階であり、その性能を引出すためには様々な技術的な課題がある。処理速度の飛躍的な向上は、学術、応用面において大きな波及効果をもたらすものと期待される。

## 2. 研究の目的

本研究では、測定器シミュレーションでの GPU の利用にいち早く注目し、GPU の超並列演算性能を引出す新しい計算手法を開発する。Geant4 の処理速度を数十倍へと飛躍的に向上させることを目的とする。GPU の演算性能は非常に高いが、CPU の計算に比べてプログラミングの自由度がない。すなわち、アプリケーションに対して積極的にアルゴリズムの改良を施し、GPU の性能を生かすプログラミングモデルを開発する。超並列計算に対応した Geant4 のデータモデル、粒子輸送アルゴリズムの開発を行い、測定器シミュレーションの高速化を実現する。アプリケーションの性能を評価し、測定器シミュレーションにおける GPU コンピューティング技術の適用性を示す。

## 3. 研究の方法

超並列計算に対応した Geant4 の並列化を行う。GPU のハードウェアとして NVIDIA 社の Tesla GPU 演算ユニット C2070 / K20 (C2070 の後継の GPU プロセッサ) を用い、GPU のプログラミング環境である CUDA を使ってソフトウェアの開発を行う。既存のプログラムをそのまま GPU 上で走らすことはできないので、新たに GPU の高い並列処理能力を引出すためのデータモデル、アルゴリズムの開発を行う。

### (1) GPU 上での超並列粒子輸送アルゴリズムの開発

図に GPU 上でのスレッド構成を示した。256 個のスレッドブロックがあり、各ブロックは

128 個のスレッドを有す。すなわち、32,768 個のスレッドを同時に実行できることになる。この GPU の並列性能を引出すためには、一般に、SIMD (Single Instruction Multi Data) と呼ばれる複数のデータに対して、一つの処理を適応させるアルゴリズムが有効である。GPU 上で多数の粒子を同時に輸送するための超並列粒子輸送アルゴリズムを開発する。

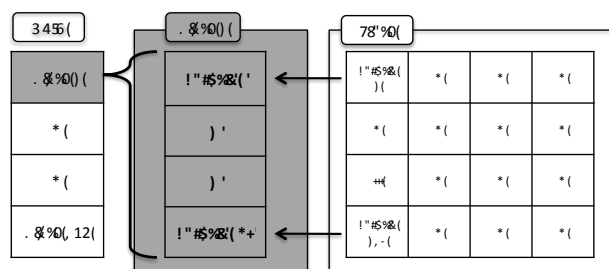


図 1 : GPU 上でのスレッドでの並列粒子輸送の処理。

### (2) GPU 上での物理過程の実装

素粒子の素過程の反応を GPU 上でシミュレーションするために、CUDA を使って、既存の Geant4 の電磁相互作用の物理過程の実装を行う。反応断面積のデータを Geant4 から取得し、GPU 上のメモリに配列データとして用意する。その断面積のデータから次に起きる反応を決定し、その後各物理過程の反応を起させる。

### (3) GPU プログラムの最適化

GPU 上でのプログラムは、多くのスレッドで同時にメモリアクセスをするために、プログラムの性能が落ちてしまうことがある。そこで、メモリアクセスの頻度が少ないアルゴリズムを採用したり、アトミックなメモリアクセスをすることで、メモリアクセスの集中による性能劣化を防ぐ必要がある。プログラムのプロファイリングを行い、処理に時間がかかっている部分を特定し、ボトルネックの改善、プログラムの最適化を行う。

### (4) アプリケーションの性能評価

実際のアプリケーションを通して、性能評価を行う。CPU と GPU 上で Geant4 のアプリケーションを作成し、それぞれのプログラムで同じ結果が得られることを確認し、GPU による高速化がどの位得られるか評価する。

#### 4. 研究成果

(1) GPU 上での超並列粒子輸送アルゴリズムの開発を行った。GPU のメモリ上に輸送する粒子、2 次粒子を蓄えとくスタックをそれぞれ用意し、各スレッドで独立に粒子輸送を行う。GPU メモリ上のデータ構造を簡素化するために、スタックは 1 次元配列で実現し、各スレッドで 40 個の 2 次粒子まで記憶できるように実装した。GPU での粒子輸送用のカーネル関数(スレッド上で動作する関数)が、スタック上の粒子データに対して同時に作用することで、並列粒子輸送を行う。

実際には、GPU 上の約 3 万 2000 個のスレッド上での並列粒子トラッキングを実現した。図 2 にはスタックサイズの変化の様子を示す。並列処理に起因するボトルネックがなく、ループ回数に応じて、順調に粒子が処理されている。



図 2: 粒子スタックサイズの変化。緑線が並列処理中の全粒子数(初期粒子+2 次粒子)、赤線が処理すべき初期粒子数、青線が、可動中のスレッド数。

(2) CUDA を使って、GPU 上で Geant4 の電磁相互作用の実装を行った。実際には、下表に挙げる物理過程を実装した。

粒子	実装した物理過程
電子、陽電子 ( $e^-/e^+$ )	エネルギー損失 多重散乱 制動輻射 対消滅(陽電子)
光子( $\gamma$ )	コンプトン散乱 光電効果 対生成

実装にあたっては、反応断面積の管理と実際の素過程の反応を記述する部分の 2 つの要素がある。反応断面積に関しては、Geant4 の断面積データを取得し、各反応に関してデータテーブル化を行い、各粒子のエネルギーの

関数として反応断面積を計算できるようにした。各粒子で次にどの素過程の反応が起きるかを、これらの断面積データをもとに確率的に決定するアルゴリズムを実装した。その後、起きる素過程のシミュレーションを行う。また、実際の Geant4 のシミュレーションと同等の性能が出るように、個々の反応ごとに正しく素過程が記述されているか検証作業を行った。一例として、図 3 に GPU でのシミュレーションによる、6MeV の  $\gamma$  線の水とのコンプトン散乱による散乱後の  $\gamma$  線と反跳電子のエネルギースペクトラムを示す。

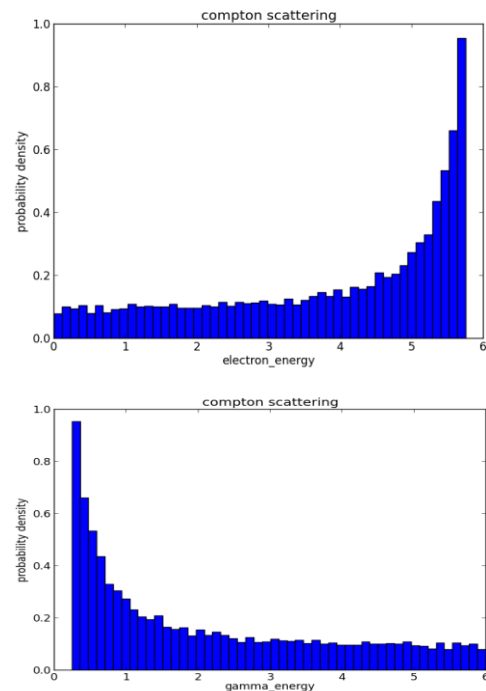


図 3 : GPU シミュレーションによる 6MeV の  $\gamma$  線の水とのコンプトン散乱による散乱後の  $\gamma$  線(上)と反跳電子(下)のエネルギースペクトラム

(3) プログラムのプロファイリングを通して、各処理での計算処理の最適化を行った。下表は、シミュレーションの各要素での処理時間のプロファイリング結果の一例を示す。

GPU シミュレーションによる各要素での処理時間

処理内容	処理時間
物理過程の計算処理	50%
エネルギー損失情報の処理	30%
粒子輸送処理	18%
プログラム管理	2%

### 各物理過程の計算処理時間の内訳

物理過程	粒子	処理時間
制動輻射	$e^-/e^+$	23%
対生成	$\gamma$	7.5%
輸送	$e^-/e^+/\gamma$	7%
光電効果	$\gamma$	7%
エネルギー損失	$e^-/e^+$	3%
コンプトン散乱	$\gamma$	1.5%
陽電子消滅	$e^+$	1%

(上表 50%の内訳)

プロファイリングの結果、エネルギー損失情報を GPU メモリ上に記憶、処理する部分に多くの処理時間がかかっていることが分った。これは、GPU 上の多数のスレッドが GPU 上のメモリに同時アクセスすることに起因する問題である。最初の実装では、並列輸送で実装した並列スタックを利用して、各ステップでのエネルギー損失情報をスタック上に 1 次的に蓄えた後、一定の頻度で Reduce 処理を行っていた。これを Atomic 演算による排他処理で行うようにすると、20%程処理速度が改善される。下表には、実際にチューニングに伴う処理速度の改善の一例を示す。GPU のハードウェアを最新の K20 に変更することで、約 15%速度向上する (K20 に特化した最適化は現時点で行っていない)。また、浮動小数点演算処理を GPU の高速関数で行うようにすると、約 5%処理速度が改善される。

### チューニングに伴う計算時間の改善

GPU	計算手法	計算時間
C2070	並列スタック	72 分
K20	並列スタック	60 分
K20	並列スタック fast math	55 分
K20	Atomic 演算	41 分

物理過程の計算処理に関しては、反応の記述の複雑さ以外にも、データテーブルへのアクセスが多いと、ボトルネックになりやすいことが分った。上表で他の過程より制動輻射に多くの処理がかかっているのは、制動輻射が内部的に 2 次元の配列データを使って反応が記述されているためである。現在のところこうしたデータテーブルは、GPU のグローバ

ルメモリ領域に配置されている。グローバルメモリ領域は、大量のメモリ空間が使える一方でメモリアクセス時間が遅いのが問題である。そこで、こうした静的なデータテーブルを、より高速なメモリ空間に配置することを検討している。実際には、GPU のテキスト領域に配置し、データの挿入処理も GPU 固有の処理を利用することで、処理速度の改善が期待できる。

(4) 評価アプリケーションとして、ボクセルジオメトリでの電磁相互作用を GPU 上で実装した。並列粒子輸送、2 次粒子の生成、電子、陽電子、光子の物理過程を CUDA を使って実装した。図 4 には、6MeV の  $\gamma$  線の水中でのエネルギー損失の分布を示した。これは、Geant4 のシミュレーションと同等のシミュレーション結果である。GPU でのシミュレーションでは、1 億イベントの処理を約 40 分で完了する。これは、Geant4 の CPU でのアプリケーションと比較して、30~40 倍程度高速化を実現し、研究計画段階で設定していた目標を実現した。更なる最適化やマルチ GPU 環境を利用することで、一層の速度向上も期待できる。

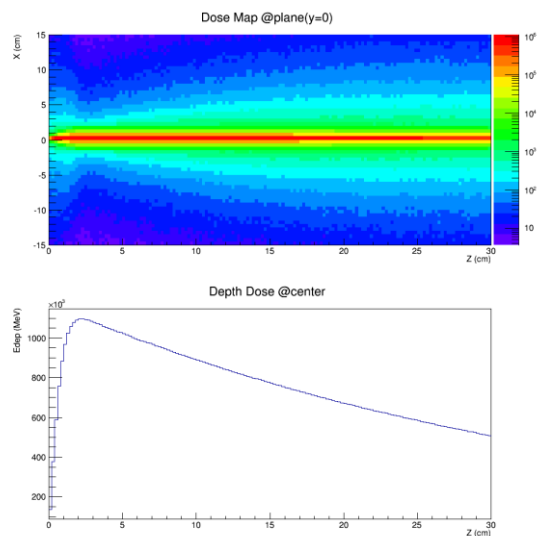


図 4 : GPU シミュレーションによる 6MeV の  $\gamma$  線の水中でのエネルギー損失。上図 : 中央付近 5mm 厚での 2 次元エネルギー損失分布。下図 : 中央 5mm x 5mm の深度エネルギー損失分布。

本研究を通して、GPU を利用することで、測定器シミュレーションの速度を大幅に向上させられることが分った。しかし、GPU の特性上、プログラミング上の制約が多くあり、現状ではどのアプリケーションでも高速化できるわけではない。例えば、ハドロン

相互作用をどのように扱うか、複雑なジオメトリの処理をどうするかなどが課題として残っている。GPU コンピューティングの技術は、まだ発展途上の段階にあり、今後も多くの新技術、機能改善が期待でき、それによって応用範囲も広がってくるであろう。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 1 件)

Koichi Murakami, CUDA-based Geant4 Monte Carlo Simulation for Radiation Therapy  
*GTC Technology Conference 2013, San Jose, USA (2013 年 3 月 20 日)*

#### 6. 研究組織

##### (1) 研究代表者

村上 晃一

大学共同利用機関法人高エネルギー加速器研究機構・計算科学センター・研究機関講師

研究者番号：10353369

##### (2) 研究分担者

なし

##### (3) 連携研究者

なし