

氏名（本籍）	富永 浩文（茨城県）
学位の種類	博士（工学）
学位記番号	甲第 226 号
学位授与の日付	平成 31 年 3 月 22 日
学位授与の要件	学位規則第 4 条第 1 項該当
学位論文題目	GPU のアーキテクチャを考慮した数値解析の高速化に関する研究
論文審査委員	(主査) 教授 前川 仁孝 (副査) 教授 宮崎 収兄 教授 富井 規雄 教授 藤田 茂 早稲田大学 教授 木村 啓二

学位論文の要旨

GPU のアーキテクチャを考慮した数値解析の高速化に関する研究

本論文は、GPU を用いた数値解析において GPU の実行効率を高めることを目的とする。防災や天気予報、製品開発などの幅広い分野で用いられる数値解析を始めとした科学技術計算は、大規模かつ精度の高いシミュレーションが求められている。高速かつ精度の高い計算できる計算機アーキテクチャとして、GPU が注目されている。GPU は、演算コアを多数搭載し、広帯域なメモリバンド幅のメモリを持つ SIMT アーキテクチャである。SIMT アーキテクチャは、複数のデータを同一の命令で処理することで高い実行性能を得られる。このため、GPU を用いる数値解析には、GPU のアーキテクチャに適した計算アルゴリズムが必要となる。

そこで、本論文では、GPU を用いた数値解析を高速化するために、アーキテクチャの特性を活かして GPU の実行効率を向上する手法を提案し、その有効性を評価する。

以下に本論文の各章の概要を述べる。本論文は全 6 章より構成される。

まず、第 1 章「序論」では、本研究における背景および従来研究について述べ、提案手法の目的や位置づけを明らかにする。

第 2 章「CUDA」では、GPU のハードウェア構成と CUDA プログラミングについて述べる。CUDA のプログラミングモデルは、スレッド階層とメモリ階層から構成される。このため、GPU の実行効率を高めるためには、CUDA の階層的構造に合わせてプログラムの局所性を抽出する必要がある。

第 3 章「メモリアクセスの局所性の向上」では、レジスタやシェアードメモリなどの階層的な

メモリを用いる最適化手法について提案する。GPU は、多くのスレッドが同時に起動できるように多くのレジスタを搭載する。このため、大量のスレッドを同時に実行でき、メモリアクセスのオーバーヘッドを隠蔽することができる。メモリアクセスのオーバーヘッドの削減には、シェアードメモリを利用する手法があり、高い効果が得られている。メモリアクセスのオーバーヘッドをより削減するためには、GPU のもつ多くのレジスタを効率的に利用することで計算をさらに高速化できる。本章では、格子ボルツマン法を題材として、局所性の高いデータをレジスタに保持することで高速化する手法を提案する。本提案手法により、テンポラルブロッキングを用いた手法と用いない手法で、最大約 7.36 倍の高速化が確認できた。

第 4 章「並列性の抽出による高速化」では、GPU の実行単位であるワーブに最適化する形で並列性を抽出する手法を提案する。命令の実行単位は、ワーブと呼ばれる 32 スレッドのまとまりで命令を実行する。このため、異なる演算が同一のベクトル命令に抽出されると、実行効率が低下する。そこで、本章では、並列性の低い問題であるランダムスパース方程式求解の LU 分解法を題材とし、命令の並べ替えを行うことでワーブの実行効率を最大化する手法を提案する。本章の提案手法により、CUDA 向けの数値計算ライブラリである CULA の LU 分解ルーチンを利用した手法にくらべ提案手法は、最大約 238 倍の高速化が確認できた。

第 5 章「並列度に応じたハイブリッド並列化手法による高速化」では、CPU と GPU で物理メモリを共有するアーキテクチャを利用し、並列度に応じて CPU と GPU を使い分けることで計算を高速化する手法を提案する。ハイブリッド並列化による計算は、CPU と GPU のそれぞれのメモリに局所性のあるデータを割当てて。一方、局所性のないデータを割当てて計算する場合は、バスを介したデータ転送が必要であり、性能が低下する。バスを用いないアーキテクチャとして、単一の物理メモリを共有して利用できるヘテロジニアスマルチコアアーキテクチャがある。本アーキテクチャは、バスを介さずにデータのやり取りができる。本章の提案手法では、拡張ベクトル化 LU 分解法を題材として、ベクトル長に応じて CPU と GPU を切り替えることで、計算を高速化する手法を提案する。本提案手法により、全ての実行レベルを GPU で行う手法にくらべ、提案手法は、最大約 26 倍の高速化が確認できた。

最後に第 6 章では、提案手法と評価結果をまとめ論文全体を総括する。

審査結果の要旨

防災や天気予報、製品開発などの幅広い分野で用いられる数値解析を始めとした科学技術計算は、大規模かつ精度の高いシミュレーションが求められている。高速かつ精度の高い計算できる計算機アーキテクチャとして、GPU が注目されている。GPU は、CPU と GPU がバスで接続されており、多数の演算コアを搭載し、広帯域なメモリバンド幅のメモリを持つ SIMT (Single Instruction Multiple Thread) アーキテクチャである。SIMT アーキテクチャは、複数のデータを同一の命令

で処理することで高い実行性能を得られる。また、多くの計算スレッドを起動するために、階層的なメモリ構造を持つ。このため、GPU を用いる数値解析には、GPU のアーキテクチャに適した計算アルゴリズムが必要となる。GPU を用いた数値解析を高速化するために、NVIDIA 社製の GPU プログラミング統合開発環境である CUDA (Compute Unified Device Architecture) を用いて、GPU による高い並列性を利用できる格子ボルツマン法、および、GPU による高速化が困難な問題であるランダムスパース方程式求解を対象とした 2 つの数値計算を題材として、GPU の特徴に合わせ計算アルゴリズムを最適化する手法について述べている。

本論文は、GPU (Graphics Processor Unit) を用いた数値解析において GPU の実行効率を高めるために GPU のアーキテクチャを考慮した最適化において重要な 3 点の最適化方法の重要性について示し、それらの最適化手法の有効性を評価したものである。

本論文は全 6 章より構成される。

第 1 章「序論」では、本研究における背景および従来研究について述べた上で、提案手法の目的や位置づけを明らかにしている。

第 2 章「CUDA」では、GPU のハードウェア構成と CUDA プログラミングについて述べられており、CUDA プログラミングにおける最適化の注意点について明らかにしている。

第 3 章「メモリアクセスの局所性の向上」では、格子ボルツマン法を題材として、CUDA のスレッドブロック内のメモリ階層に着目した最適化手法が提案されている。本提案手法は、テンポラルブロッキングと呼ばれる手法を格子ボルツマン法の計算アルゴリズムに適用し、スレッドブロック内のシェアードメモリとレジスタの階層的な利用を可能にして、計算を高速化している。評価の結果、本提案手法を用いた手法は、テンポラルブロッキングを用いた手法と用いない手法で、最大約 7.36 倍の高速化に求解できることが示されている。

第 4 章「並列性の抽出による高速化」では、ランダムスパース方程式求解を題材として、CUDA の実行単位であるワープに最適化する形で並列性を抽出する手法が提案されている。本提案手法は、並列性の抽出が困難な方程式求解手法を拡張ベクトル化 LU 分解法により、高い並列性を実現するとともにワープと呼ばれる CUDA の命令実行形式の重要性を示している。ワープ内で異なる命令を実行することで CUDA の実行効率が低下することを示し、異なる演算命令が含まれないベクトル抽出方法を示し、計算を高速化している。評価の結果、異なる命令が実行される従来手法に比べて提案手法は、最大で約 1.6 倍高速化に求解できることが示されている。

第 5 章「並列度に応じたハイブリッド並列化手法による高速化」では、ランダムスパース方程式求解を題材として、並列度に応じて CPU と GPU を使い分けることで計算を高速化する手法が提案されている。本提案手法は、ランダムスパース方程式求解において、ベクトル長が低い命令が抽出される点に着目し、CPU と GPU によるハイブリッド並列化によるデータ転送のオーバーヘッドが問題になること述べている。この問題を解決できるアーキテクチャに、近年着目されている SoC (System on Chip) を用いた解決策を提示している。SoC は、CPU と GPU でメモリを共有化できるため、データ転送のコストを削減できるアーキテクチャである。この点に着目し、CUDA の命令実行の粒度に合わせた CPU/GPU 切り替え手法を提案している。評価の結果、全てのベクトル命

令を GPU で行う手法に比べて、提案手法は、最大約 26 倍の高速化に求解できることが示されている。

最後に第 6 章では、提案手法と評価結果をまとめ論文全体を総括している。

本論文は、GPU のアーキテクチャを考慮することで数値計算を高速化する手法について研究したものであり、GPU を用いた数値計算のアルゴリズムを最適化するための重要な知見を得たものとして価値があると認める。

従って、学位申請者の富永浩文は、博士（工学）の学位を得る資格があると認める。