T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	大規模並列ヘテロジニアス環境におけるメモリ構造を考慮したソーティングおよび配列アラインメントの最適化
Title(English)	Memory-Conscious Optimizations for Sorting and Sequence Alignment for Massively Parallel Heterogeneous Architectures
著者(和文)	DROZD Aleksandr
Author(English)	Aleksandr Drozd
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第9420号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:松岡　聡,遠藤　敏夫,山下　真,脇田　建,渡辺　治
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第9420号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Powered by T2R2 (Tokyo Institute Research Repository)

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	数理・計算科学　　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of Science　（理学）
学生氏名： Student's Name	Aleksandr Drozd	指導教員（主）： Academic Advisor(main)	教授　松岡　聡
		指導教員（副）： Academic Advisor(sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

本学位論文は「Memory-Conscious Optimizations for Sorting and Sequence Alignment for Massively Parallel Heterogeneous Architectures」と題し，最新の計算機環境において大規模データを対象とする計算アルゴリズムの実行性能を向上させる手法について述べている．中でも特に重要なソートと配列アラインメントの２つの問題に着目している．本論文は以下のような英文５章で構成されている．

第一章「Introduction」では，本研究の対象とする背景，問題提起及びそれに対する主要な成果について総括的に述べるとともに，本論文の構成を示している．

第二章「Background」では，本研究に関連する背景をより詳細に述べている．まず ヘテロ型計算環境とそのプログラミングモデルについて，具体的に GPU による並列プログラミング手法やスーパーコンピュータ TSUBAME 2.5 を挙げて詳細に説明している．また，大規模ヘテロ型計算機上で高性能計算を実現するための課題について述べている．具体的には，本論文で対象としている配列アライメント及びソートの高速化のための，階層型メモリの効率的な利用，ホストメモリとデバイスメモリ間の通信や負荷分散について述べている．

第三章「Sequence Alignment」では，配列アラインメントの問題に関し，ヘテロ型計算環境における既存の手法の問題を分析し，新たな手法について述べている．まずゲノム計算分野における問題の概略について解説し，問題を厳密に定義している．そしてデータ構造まで詳細に説明し，選択されたアルゴリズムを GPU アーキテクチャで実行する場合の効率という観点での分析を行っている．特にボトルネックとなりやすい GPU とホスト間のデータ転送のコスト削減に注力している．またデータ構造を見直し，コンパクト化することで配列アラインメント処理の性能向上を実現している．Burrows-Wheeler 変換による suffix array 法の利用及び探索アルゴリズムの改良によって既存の高速な GPU 実装である MummerGPU++の 3~4 倍の計算速度を達成している．データサイズが GPU のデバイスメモリ容量を超える場合についても性能モデルを構築し，性能が最高となるように一度にデバイス上で処理するデータのサイズをモデルから算出することに成功しており，また複数 GPU の利用にも対応している．

第四章「String Sorting」では，既存のソートアルゴリズムをデータ型など特定の計算機アーキテクチャに対する適応性の面で分類している．文字列データのソートに適用可能なソートアルゴリズムの中で，GPU に適していると判断された most significant digit による基数ソートに焦点をしぼっている．文字列データを対象とするヘテロな計算機における，ソートの並列化は世界で初の試みであり，並列実行時のスケーラビリティを改善するため異なる並列化手法を CPU と GPU で２段階に組み合わせた手法を提案している．GPU 用の実装ではホストからデバイスへのデータ転送がボトルネックとなっていることが判明したため，通信量削減手法によりこの問題を解決している．

第五章「Discussion and Conclusions」では，提案手法の実装経験や性能評価から得られた知見をまとめている．それらは今回対象とした分野だけでなく，スーパーコンピュータプログラミング一般に適用可能である．この研究成果の計算生物学や文字列データ解析等に対する寄与は大きく，また大規模データを対象とした各種アルゴリズムをアクセラレータ型計算機システムに対して最適化する場合の指標となるものである．

以上のように，本論文は大規模並列環境を対象とした配列アラインメントやソート処理の高速化手法を提案し，その有効性を証明しており，理学上貢献するところ大である．よって本論文は博士（理学）の学位論文として十分価値があるものと認める．

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	数理・計算科学　　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of Science　（理学）
学生氏名： Student's Name	Aleksandr Drozd	指導教員（主）： Academic Advisor(main)	教授　松岡　聡
		指導教員（副）： Academic Advisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

Thesis addresses the issue of improving performance of data-intensive algorithms on modern computing architectures. It is composed of 5 chapters. First, brief introduction to the whole work is given. Chapter 2 provides background information about HPC and parallel programming, particularly on CUDA and GPU architecture, overview known algorithm optimization strategies, and provide some examples of their successful implementations.

Chapter 3 addresses the problem of sequence alignment. We start by providing an introduction to the problem domain and a formal definition of the problem. For the sequence alignment we adapted Burrows Wheeler transform based index to reduce overall memory footprint on GPUs. We describe in detail our data structure and analyse the chosen algorithm in terms of its efficiency on GPU architecture. Particular attention is given to the problem of reduction of costs of moving data between host and GPU memories. A mathematical model of computation and communication costs was developed to optimize memory partitioning for index and queries. Performance evaluation shows more than ten-fold performance gain per device.

Chapter 4 presents the classification of sorting algorithms with respect to the data types and fitness of the algorithms to particular architectures. We focus on the most-significant digit radix sort, one of the few algorithms that are applicable to sorting of string data. We present CPU and GPU implementations of the algorithm using different parallelization strategies on different execution stages to optimize the use of system resources and workload balance. Our solution achieves sorting rates up to 70 million keys per second sorting throughput with good scalability. This chapter also discusses our performance model: load balancing, scalability, and communication expenses.

Chapter 5 concludes outlines the main findings from implementation experience and performance evaluation of our solutions. We and also discuss the implications of our findings for computational biology and string data analysis and suggest some directions for future work.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).