

令和 4 年 8 月 31 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K11336

研究課題名(和文) 超並列において高スケーラビリティを実現するステンシル計算・通信モデルの開発

研究課題名(英文) Development of stencil calculation and communication model to achieve high scalability on massively parallel computation

研究代表者

深沢 圭一郎 (FUKAZAWA, Keiichiro)

京都大学・学術情報メディアセンター・准教授

研究者番号：50377868

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究では、エクサスケール環境においてスケーラビリティ減衰が無いステンシル計算・通信モデルの開発、及びそこで利用されるHalo通信関数の開発を行うことを目的とした。まずステンシルシミュレーションにおいて、「計算」と「通信が必要な計算と通信」にスレッドを分けるモデルを開発した。これにより、通信が終わったことを知るための同期が必要無く、並列性能劣化を回避することができた。次に、そこで利用された通信モデルを関数群(Halo関数)にまとめ、他のアプリケーションでも容易に利用可能とした。これらの性能を2000ノード利用した環境で測定を行い、高いスケーラビリティを確認した。

研究成果の学術的意義や社会的意義

近年の計算機はプロセッサ単体の性能を周波数向上が難しいことから上がられず、目に-コア化や並列ノード数増加により性能向上を達成している。この様な中で並列計算時にスケーラビリティが低下すると、並列数増加分の性能向上が得られない。本研究の成果により高並列時にもスケーラビリティが減少しない手法が開発され、近年の計算機をより効率良く利用することが可能となる。また、ステンシル計算はある空間を扱う計算で用いられる計算であり、スパコンを利用するような大規模計算アプリケーションの大部分を占めるため、幅広い領域へ貢献することが推測され、学術的意義が深い。

研究成果の概要(英文)：The purpose of this study is to develop a stencil calculation and communication model without scalability attenuation in exascale computing environment, and to develop the Halo communication function used there. First, in the stencil simulation, we developed a model that divides threads into "calculation" and "calculation and communication that require communication (including pack / unpack of communication data)". As a result, there is no need for synchronization to know when the communication has ended, and deterioration of parallel performance can be avoided. Then, we summarized a Halo function group from the model so that it could be easily used in other applications. These performances were measured in an environment using 2000 nodes, and high scalability was confirmed.

研究分野：高性能計算

キーワード：並列計算 高性能計算 MHDシミュレーション

### 1. 研究開始当初の背景

現在、CPU は電力問題から周波数の向上は望めず、その性能向上は CPU 当たりのコア数を増やすこと(メニーコア化)と 1 クロック当たりに同時演算可能な数を増やすことによって行われている。また、CPU 単体での劇的な性能向上が見られない中、いわゆるスーパーコンピュータ(スパコン)を始めとする HPC(High performance computing)システムは大多数の計算機をネットワーク接続(計算ノード間接続)し、その全体性能を向上させている。この

傾向は京コンピュータに代表されるペタスケール計算機システムから、現在開発が進んでいるエクサスケール計算機システムにおいても変わることは無い。そのため、今後、HPC システムを利用するアプリケーションは、コア数の増加、接続される計算機の増加により、劇的に並列数が増加した中での計算が余儀なくされる。

計算機の理論演算性能は、このような超並列ネットワーク接続、メニーコア化により高くなるが、実際に計算機を利用する場合には、計算ノード間や CPU コア間の通信、それに伴う同期といったオーバーヘッドがかかってくるため、理論演算性能が発揮されることは無い。実際、格子計算系(ステンシル計算)モデルでは、並列化を行う場合に空間格子を並列計算プロセス数で分割する領域分割法を導入することが一般的であり、分割に伴いデータの送受信が発生する。このような超並列環境下で、計算格子サイズは変えず、並列数を増やすストロングスケーリング(並列数が増えると、プロセス当たりの計算量は減少する場合)では、いわゆるアムダールの法則(通信時間が計算時間を超えると性能は上がらない)に従ってしまい、並列数を上げて計算性能があまり上昇しない場合も多い。また、並列数を上げてプロセスあたりの計算量を変更しないウイークスケーリングにおいても、計算時間と通信時間の比率が変わらないために、高い並列化効率(並列化数の分性能が向上しているかをみる割合)を維持していたが、超並列環境ではプロセス間同期の問題から、並列化効率(スケーラビリティ)が悪くなってきていることが申請者の研究により報告されている。その中では、ペタスケールである京コンピュータを利用する場合においても、高並列計算(3 万プロセス並列)では、スケーラビリティが劣化することが分かっている(図 1)。また、メニーコア CPU では、1 コア当たりの性能が通常の CPU に比べて高くないため、通信やそれに伴う処理が想定よりも時間がかかってしまうことも分かっている。

エクサスケール計算機システムは、2020 年過ぎに完成が計画されているが、数百万ノード接続の構成と想定されている。実利用上ノード内複数プロセス並列で計算が行われることも考えると、並列プロセス数は数千万のオーダーになる。図 1 の結果から、この並列計算環境でのウイークスケーリング性能を見積もったところ、20%以上の並列性能劣化となった。エクサスケールにおいて、現在の実行計算効率で計算できた場合、20%の性能劣化は、数ペタフロップス(京コンピュータ 1 台分程度)の損失となり、アプリケーションの計算性能の問題だけではなく、経済的コストの面でも、システム運用面からも非常に大きな問題となることが容易に考えられる。また、エクサスケールでは、高性能メモリを利用する関係上、ノード当たりのメモリ量の増加が見込めず、ある程度のストロングスケーリングでの計算が強いられるため、更なる並列性能劣化が起こると危惧される。

このように将来のエクサスケール計算環境において、これまでスパコンで利用されてきた研究者独自開発のアプリケーション(ステンシル計算)は、これまでと同様にスケールアップして動作ができるのかと

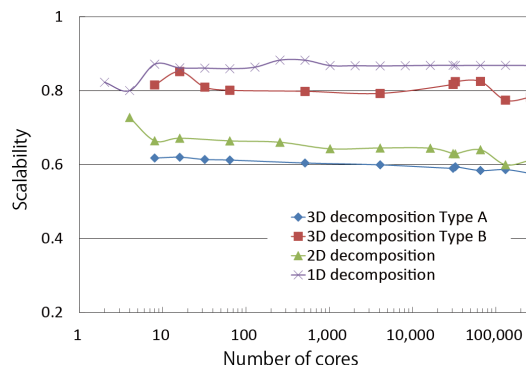


図 1 電磁流体コードの京コンピュータにおける並列性能

ということが直近において課題となりつつある。

## 2. 研究の目的

本研究では、数値計算としては大部分を占めるステンシル計算に注目し、今後の超並列計算、エクサスケールコンピューティング時代において高スケーラビリティを達成することができる計算・通信フレームワークを開発することを目的とする。

これまでの研究では高スケーラビリティを達成するために、通信時間の隠蔽が考えられており、通信専用スレッドにより通信を計算に隠蔽する手法が多かった。しかしながら、通信終了を知らせる同期処理が全スレッド間で必要であり、場合によってはプロセス間でも必要であった。前述のようにこの同期はエクサスケールコンピューティングでは、非常に大きな問題となるが、その対応について今のところ確立された技術は無い。本研究では、通信の隠蔽を行った上で、初めてスレッド間同期処理を取り除いた通信と計算モデルを開発することが最も独自性があり創造性が認められるという点であると言える。更に、通信の準備のため行う通信されるデータのパック・アンパックもスケーラビリティに大きく影響するが、本研究で開発するモデルでは、この動作も隠蔽可能なため、スケーラビリティ低下を抑えることが可能である。

エクサスケール級計算機であるポスト京向けアプリケーションとしては、重点課題や萌芽的課題が選ばれており、そこに含まれる有名アプリケーションはこれらの問題に早くから対処し、性能向上を目指している。しかしながら、これら課題に選ばれているアプリケーションは、例えば京を利用しているアプリケーションの一部に過ぎず、大部分のアプリケーションはその恩恵を受けることが難しい状況である。一方で、本研究で開発する通信と計算のモデルは、ステンシル計算で一般的な並列化手法である領域分割法を利用した並列化を行っている場合には、ほぼすべてのコードに適応できるシンプルな構造である。また、導入方法も MPI でプロセス並列されていれば、その部分を書き換え、追加を行うことで導入が基本的に可能となる計画のため、既存の研究者個人開発アプリケーションに導入が容易である。加えて、研究対象とするステンシル計算はある空間を扱う計算で用いられる計算であり、スパコンを利用するような大規模計算アプリケーションの大部分を占めるため、幅広い領域へ貢献することが推測され、学術的意義が深い。

本研究で開発する通信と計算のモデルでは、同期を取り除くために、「主な計算部分」と「通信が関わる計算部分」を分離する。これにより、アプリケーションの計算性能に関わる「主な計算部分」に対して変更すること無く、「通信が関わる計算部分」だけに変更、最適化を加えることで、新しい通信ライブラリや通信モデルを導入することができ、生産性、メンテナンス性が良い。そのため、本研究で開発されるモデルがエクサスケールに向けたステンシル系アプリケーションの創造に繋がると同時に標準的なシミュレーションモデルになっていくと考えられ、社会実装につながるものと期待される。

## 3. 研究の方法

エクサスケール級計算において、並列性能が劣化するという問題に対し、現在ハードウェア側からもミドルウェア側からも研究が盛んに行われており、同様にアプリケーション側からも、通信と計算を並列に実行して、通信時間を隠蔽する技術が開発されている。しかしながら、この技術では通信時間を隠蔽するだけとなっており、通信自体の最適化は行われず、また明示的に同期を取る必要があり、エクサスケールでの並列性能は保証されない。そこで、本研究では、下記の二つの研究を期間内に実施した。

### (A) スケーラビリティを高く維持できるステンシル計算・通信モデルの開発

まず、ステンシルシミュレーションにおいて、「計算」と「通信が必要な計算と通信(通信データのパック

ク・アンパックも含む)」にスレッドを分ける。このように分割すると、通信が終わったことを知るための同期の必要無く、並列性能劣化を回避することができる。ステンシル計算で通信が必要な場所は、主に領域分割に伴う通信である。図2に示すように、あるプロセスで計算を行う場合に、周辺プロセスが持つ値が必要となるために、その値を転送する必要がある。この転送されるデータを「袖 (Halo) 領域」と呼んでいる。この Halo 領域を利用する計算と Halo 通信を一つのスレッド (Halo スレッド) に割り当て、その他の計算を別のスレッドに割り当てる計算モデルが開発する具体的なモデルである。簡易的な実験をすでに行っており、このモデル自体の有用性は初期確認ができています。しかしながら、簡易計算モデルを利用した実験結果のため、実際に研究に利用されている申請者開発の3次元電磁流体シミュレーションコード、連携研究者開発の5次元 Vlasov シミュレーションコードに、対応できるモデルを研究分担者と共同で開発・実装し、その有用性を明らかにすることで、最終的には、一般的ステンシル計算に実装ができるフレームワークを開発する。

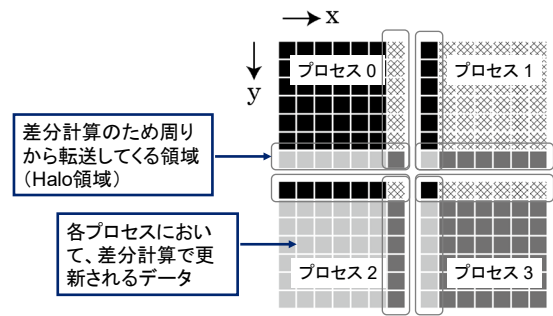


図2 並列ステンシル計算における Halo 領域

### (B) 効率的な Halo 通信モデルの開発

Halo 通信自体も通信の順序性やデータのパック・アンパック、点や線、面を通信する場合など、非効率な要素や複雑な要素が含まれるため、それらを容易に効率的に行うことができる Halo 通信関数群を作成する。例えば、現在はアプリケーション側で通信するプロセスの位置関係やデータ量など指示しているため、複雑なコードになっており、通信の効率化が難しい。この Halo 通信関数群をまずは MPI を利用して作成し、最終的には、現在開発されているエクサスケール向け通信ライブラリを利用した通信関数群とする。この関数開発は、通信ライブラリに造詣のある分担研究者と共同で行い、特に通信効率化の部分に注力する。この構成であれば、Halo 関数を利用しているアプリケーションは、Halo 関数の性能向上に伴い、コードを変更することなく、その恩恵を受けることができる。

### 4. 研究成果

まず、スケーラビリティを高く維持できるステンシル計算・通信モデルの開発をおこなった。ステンシルシミュレーションにおいて、「計算」と「通信が必要な計算と通信 (通信データのパック・アンパックも含む)」にスレッドを分けるモデルを開発した。このように分割すると、通信が終わったことを知るための同期の必要無く、並列性能劣化を回避することができる。ステンシル計算で通信が必要な場所は、主に領域分割に伴う通信である。あるプロセスで計算を行う場合に、周辺プロセスが持つ値が必要となるため、その値「袖 (Halo) 領域」を転送する必要がある。この Halo 領域を利用する計算と Halo 通信を一つのスレッド (Halo スレッド) に割り当て、その他の計算を別のスレッドに割り当てる計算モデルとなっている。このモデルを、研究代表者が開発している電磁流体力学 (MHD) コードに適用し、その性能

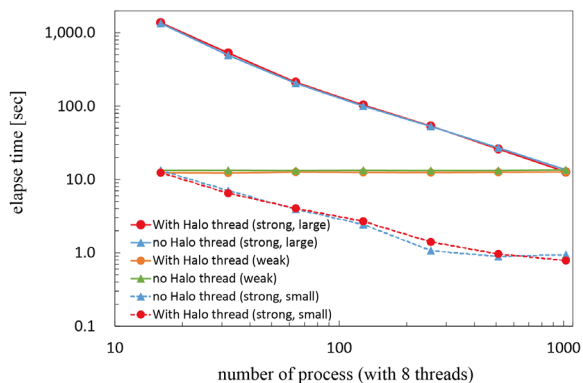


図3 FX10 を利用したストロング・ウィークスケールでの Halo スレッドの性能

を評価した。その結果、ストロングスケールにおいて、高い並列度においても、並列化効率が落ちにくく、性能向上が確認できた（図3参照）。従来の並列計算モデルでは、同並列度において、計算性能の向上が見えない（計算性能が飽和している）ことも確認できた。

次に、開発したステンシルシミュレーションにおいて、「計算」と「通信が必要な計算と通信（通信データのパック・アンパックも含む）」にスレッドを分けるモデルから、そこで利用された通信モデルを

関数群（Halo 関数）にまとめ、他のアプリケーションでも容易に利用可能とした。この Halo 関数は Halo 通信とそれに必要なパック・アンパック、また、通信と計算のオーバーラップを容易に実装できる中身になっている。現在は MPI で実装されているが、必要があれば他の通信ライブラリに置き換えることも可能であり、一方で Halo 関数を利用しているプログラム自体には変更を行うこと無く、そのライブラリ変更の恩恵を受けることができる。この Halo 関数を用いて、スケーラビリティの減衰が抑えられる計算・通信モデルの開発を進め、その効果を調べた。計算量や並列数により、効果は異なるが、通常の MPI による Halo 通信に比べ良い性能を得ることができた（図4）。

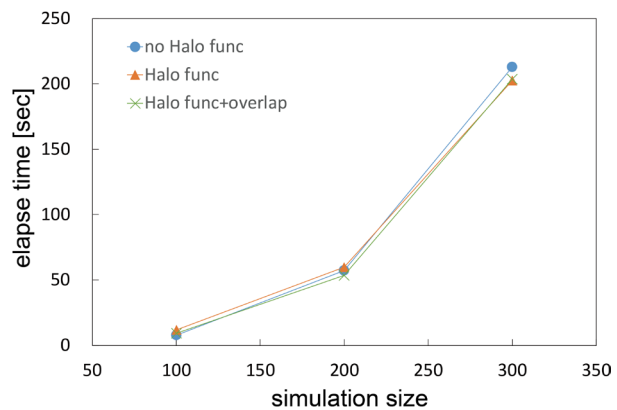


図4 XC30における Halo 関数を利用しない場合、Halo 関数を導入した場合とオーバーラップをさせた場合の測定結果

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Fukazawa Keiichiro, Katoh Yuto, Nanri Takeshi, Miyake Yohei	4. 巻 -
2. 論文標題 Application of Cross-Reference Framework CoToCoA to Macro- and Micro-Scale Simulations of Planetary Magnetospheres	5. 発行年 2019年
3. 雑誌名 Seventh International Symposium on Computing and Networking Workshops (CANDARW)	6. 最初と最後の頁 121-124
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CANDARW.2019.00029	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Fukazawa, K., T. Soga, T. Umeda, T. Nanri	4. 巻 1
2. 論文標題 Performance Evaluation and Optimization of MagnetoHydroDynamic Simulation for Planetary Magnetosphere with Xeon Phi KNL	5. 発行年 2018年
3. 雑誌名 Advances in Parallel Computing	6. 最初と最後の頁 178 - 187
掲載論文のDOI（デジタルオブジェクト識別子） 10.3233/978-1-61499-843-3-178	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 K. Fukazawa, M. Ueda, Y. Inadomi, M. Aoyagi, T. Umeda and K. Inoue	4. 巻 1
2. 論文標題 Performance Analysis of CPU and DRAM Power Constrained Systems with Magnetohydrodynamic Simulation Code	5. 発行年 2018年
3. 雑誌名 HPCC	6. 最初と最後の頁 626-631
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/HPCC/SmartCity/DSS.2018.00113	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計11件（うち招待講演 1件／うち国際学会 4件）

1. 発表者名 深沢圭一郎、三吉郁夫
2. 発表標題 惑星磁気圏MHDシミュレーションコードによるA64FXプロセッサ（FX700）の性能評価
3. 学会等名 情報処理学会第175回HPC研究会
4. 発表年 2020年

1. 発表者名 深沢圭一郎、木村智樹、徳永旭将、中野慎也
2. 発表標題 宇宙プラズマ現象予測モデル開発に向けた機械学習・数値シミュレーション・観測による学習データの整備
3. 学会等名 地球電磁気・地球惑星圏学会 第148回総会及び講演会
4. 発表年 2020年

1. 発表者名 K. Fukazawa
2. 発表標題 Research of applying the machine learning to space plasma physics with observation and numerical simulation data
3. 学会等名 LHAM Keynote in CANDAR2020 (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 K. Fukazawa, Y. Kato, T. Nanri, Y. Miyake
2. 発表標題 Development of cross-reference framework CoToCoA and application to macro- and micro-scale simulations in the space plasma
3. 学会等名 AGU Fall Meeting 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 深沢圭一郎
2. 発表標題 惑星磁気圏MHDシミュレーションコードによるThunderX2 ARMプロセッサの性能評価
3. 学会等名 第172回ハイパフォーマンスコンピューティング研究発表会
4. 発表年 2019年

1. 発表者名 深沢圭一郎、加藤雄人、三宅洋平、南里豪志
2. 発表標題 Development of cross-reference framework for macro- and micro-scale simulations of the magnetosphere
3. 学会等名 地球電磁気・地球惑星圏学会 第146回総会及び講演会
4. 発表年 2019年

1. 発表者名 K. Fukazawa, Y. Katoh, T. Nanri, Y. Miyake
2. 発表標題 Development of cross-reference numerical simulation framework: macro- and micro-scale simulations of the magnetosphere
3. 学会等名 Conference on Magnetospheres of the Outer Planets 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 K. Fukazawa, Y. Katoh, T. Nanri, Y. Miyake
2. 発表標題 Development of Coupling Framework for Macro and Micro Scale Simulations of the Magnetosphere
3. 学会等名 AOGS 16th Annual Meeting
4. 発表年 2019年

1. 発表者名 深沢圭一郎、梅田隆行
2. 発表標題 MHD, VlasovシミュレーションにおけるHalo通信の最適化
3. 学会等名 STEシミュレーション研究会 - プラズマ-大気複合システムのシミュレーション研究 -
4. 発表年 2018年



1. 発表者名 深沢圭一郎、南里豪志、本田宏明
2. 発表標題 スーパーコンピュータシステムIT0におけるMHDシミュレーションコードの計算性能・消費電力評価
3. 学会等名 第166回ハイパフォーマンスコンピューティング研究発表会
4. 発表年 2018年

1. 発表者名 K. Fukazawa, M. UEDA, Y. INADOMI, M. Aoyagi, T. UMEDA, K. INOUE
2. 発表標題 Performance Analysis of CPU and DRAM Power Constrained Systems with Magnetohydrodynamic Simulation Code
3. 学会等名 International Conferences on High Performance Computing and Communications 2018 (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	南里 豪志  (NANRI Takeshi)  (70284578)	九州大学・情報基盤研究開発センター・准教授    (17102)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------