様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １６ 日現在

機関番号：３４５０６
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１１８４１

研究課題名（和文）メニコア大規模クラスタ向け分散データ管理ライブラリおよびタスク管理機構との融合

研究課題名（英文）Distributed Data Management Library for Large-Scale Many-Core Clusters and its Integration with Dynamic Load Balancers

研究代表者
鎌田　十三郎（Kamada, Tomio）

甲南大学・知能情報学部・准教授

研究者番号：２０３０４１３１

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：メニーコア大規模クラスタのための分散データ管理ライブラリの開発および動的タスク管理機構との融合を目指した研究をおこなった.
具体的には(1) メニーコア環境における動的負荷分散の効率化のためのタスク粒度自動調整機能、(2) メニーコア環境に対応した高並列データアクセスが可能な分散集合ライブラリの拡充をおこない、加えて(3) 動的負荷分散機構と分散データ管理機構の融合にむけ (3a) 分散配列ライブラリに対する要素の配置を伴う動的負荷分散機構の導入および (3b) 通信と引き続く計算処理の依存関係を容易に記述可能な分散セル集合ライブラリの研究をおこなった。

研究成果の学術的意義や社会的意義
今後、スーパーコンピュータの用途が広がりや計算の高知能化により、並列プログラムの不規則化が予想される。例えば、状況に応じて大規模な計算をする知的なエージェントをシミュレートする場合、負荷状況に応じた計算資源の再割り当てが必要になる。一方で、現在のスーパーコンピュータでは、メニーコアプロセッサが一般化するなど、より大規模化・複雑化が進んでいる。
ノード間にまたがるデータ・タスク配置管理の今後のさらなる複雑化に対応するため、本研究では、メニーコア環境における要素の再配置可能な大規模分散データ管理ライブラリを開発するとともに、自動負荷分散機構との融合に向けた研究をおこなった。

研究成果の概要（英文）：This research aims to provide distributed collection libraries for many-core large-scale clusters and enables dynamic load-balancing over them. We developed (1) a self-adjusting task granularity mechanism for our global load balancer library to avoid contention on many‐core clusters and (2) a series of relocatable distributed collections featuring inter/intra-node parallelism.
For integrating load-balancer and distributed collections, we developed (3a) a global load balancer for distributed arrays involving range-based element relocation. In addition, we developed (3b) a distributed cell set that allows the easy description of communication/computation overlapping and relationships between inter-node communication and its dependent computations.

研究分野：並列分散処理 プログラミング言語

キーワード：動的負荷分散　分散集合ライブラリ　メニーコアクラスタ

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

社会事象シミュレーションなどの分野では，エージェントの知能化が進むにつれて，しばしば複雑な計算が必要とされる．例えば，金融市場シミュレーションにおいては，各種アセットからなるポートフォリオを再構成する際，複雑な計算を要する．このような計算は，例えばアセットの価格変動が大きくなった段階で必要となるものであり，シミュレーション実行基盤には，突発的な負荷変動への対応が求められる．
一方で，高並列計算環境は計算ノード数の増大と同時に，メニーコア CPU や GPGPU の導入が進んでおり，大規模化・階層化が進展しており，その資源管理は複雑化してきた．例えば、京コンピュータでは８コアだった CPU が富岳においては４８コアまで増加している。このため、通信ボトルネックの発生は膨大な計算資源を無駄にすることに直結し、またノード間の通信・タスクのスケジューリングだけでなく、ノード内の並列処理にも細心の注意が必要となるなど、不規則性の高いプログラムをスーパーコンピュータ上で実行するうえでの課題が増大している。

２．研究の目的

本研究の目的は，不規則性の強い問題を対象に，プログラマが容易に大規模分散データを作成し，計算状況に応じて容易に適切な計算ノードにデータを配置できる環境を実現することにある．データの論理的な構造とデータ配置を分離し，計算状況に応じてデータ分散を動的に変更したり、要素データのキャッシュを他ノードにキャッシュしたりすることができる．
本研究では，(1) メニーコア環境に対応した動的負荷機構機構の開発・拡充を進め，(2) 分散集合ライブラリのメニーコア対応のため，ノード内の多数のスレッドからの並列データ登録・参照が可能なライブラリの整備やデータの階層的配置手法を実現する．加えて，(3) 分散データ管理部と動的負荷分散機構を融合することで，突発的な負荷変動に応じて関連データの再配置も可能とする．単に負荷分散に応じたデータ再配置をおこなうだけでなく，大規模並列処理を容易に記述するためのプログラミングモデルを検討する．

３．研究の方法

研究を進めるにあたっては，上記の研究 (1)，(2)，(3) を並行して進める．早期に，システムのプロトタイプを実現し，その後，段階的にメニーコア対応分散データ管理ライブラリの拡充およびプログラミングモデルの検討をおこなう．まず，連想配列や分散配列などのフラット型データに対して，システム実現・アプリケーション開発およびプログラミングモデルの改良を進め，その後，木構造データ構造などの拡充を図る．
システムの開発にあたっては，我々のグループで開発してきた分散集合ライブラリをベースとする。ただし、高性能分野向けの X10 プログラミング言語で記述されていたため、多くのデータ構造ライブラリを有し広く利用者のいる Java 言語と，その分散ライブラリの一つである APGAS ライブラリをベースに再実装する。
研究は、代表者の鎌田および研究室所属学生であった Patrick Finnerty を中心に進めており、2022 年 2 月に Finnerty が神戸大の助教に就任した際は研究分担者として参加した。

４．研究成果

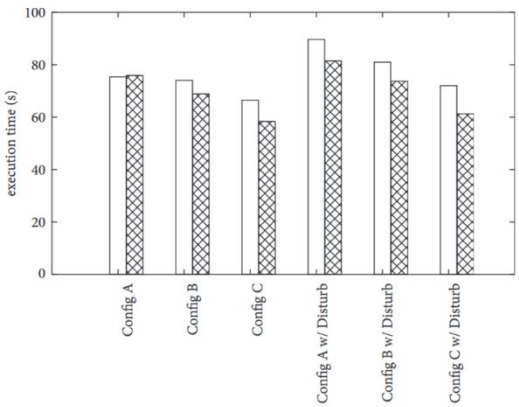
(1) メニーコア環境における動的負荷分散の効率化のためのタスク粒度自動調整機能

メニーコア環境を対象とした動的負荷分散機構に対し、タスク粒度と実行効率の関係を調査し、加えてノード内負荷分散時のメモリアクセス競合およびノード間負荷分散効率を考慮したタスク粒度管理機構を提案した。本研究は研究期間開始前からおこなっていたもので、2021 年に雑誌論文[1]にて成果発表をおこなうとともに動的負荷分散機構のシステムを公開した。ベンチマークには従来用いていた UTS に加え、N-Queen，Pentomino，TSP を加え、また大規模実験については，68 コア NUMA CPU を持つ Oakforest-PACS を用いた評価をおこなった．

(2) メニーコア環境に対応した高並列データアクセスが可能な分散集合ライブラリの拡充

2020 年からメニーコア環境において複数スレッドからのデータ挿入およびその後のノード間データ再配置を考慮した分散連想配列の実装法の研究をすすめ、また各種機能を実装したうえでソフトウェア公開を開始した．2021 年には，分子動力学および人工市場シミュレーションなどへの対応のため，２次元配列や，複数配列を対象とした多重ループ処理，連想配列に対する

ライブラリ拡充をおこなった．この研究では、システムによる自動負荷分散はおこなわれていないが、プログラマが明示的にデータ分散を変更することで負荷の平坦化を図ることは可能であり、性能評価でも明示的負荷平坦化の効果を測定した。その際、実行プログラム中の負荷変動ではなく、実行環境が提供する計算資源の変動を想定した明示的負荷平坦化実験などもおこなった。右図は人工市場シミュレーションにおいて明示的負荷平坦化(網掛)を施した際の効果を示したものである。ここまでの成果は 2022 年発行の雑誌論文[2]にて研究発表している。またシステムおよびベンチマークプログラム群も公開済みである。また、2022 年度から空間粒子シミュレーションなどを想定とした分散セル集合ライブラリのデザイン・研究開発をおこなったが、これは（3b）として後ほど解説する。

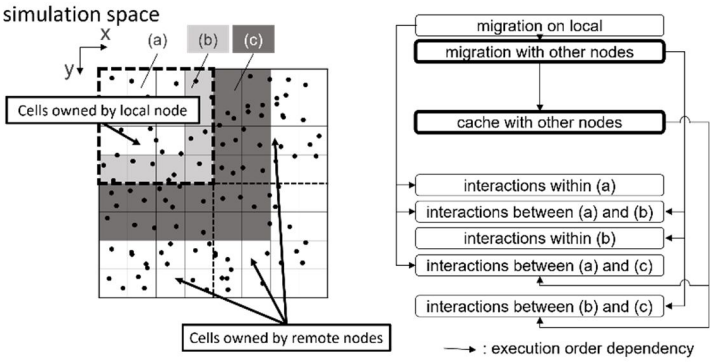


(3a) 分散配列ライブラリに対する要素の配置を伴う動的負荷分散機構の導入

2020 年度から配列型データ構造 DistChunkedList を対象に，分散集合と動的負荷分散機構との融合をすすめた．また、現実のシミュレーションでは、一連の計算フェーズを繰り返すことが多いため，複数の計算コンポーネントから構成されるようなプログラムについても，詳細なデータ配置を記述しない形で制御フローを素直に記述できるようなプログラミングモデルの研究を開始した．最終的には DistChunkedList の全要素に対し、オーナーコンピューティングタイプの並列計算をおこなう forEach 命令に要素の再配置をともなうの動的負荷分散機構を導入し、K-means や人工市場シミュレーションといったアプリケーションへの応用が可能となった。(2)で述べた明示的負荷平坦化とは違い、システムは各ノードで並列計算をおこない、早期にタスクを実行し終えたノードが他ノードから負荷を譲り受ける形で負荷分散をおこなう。本成果については 2022 年の学会発表[1]および 2023 年の雑誌論文[3]にて発表済みである。

(3b) 通信と引き続く計算処理の依存関係を容易に記述可能な分散セル集合ライブラリの研究。

2022 年度から空間粒子シミュレーションなどを想定とした分散セル集合ライブラリのデザイン・研究開発をおこなった。各セルが部分空間を表し、その空間内の粒子を管理する。各セルをノード群に割り当て・再配置可能なだけでなく、粒子のセル間移動やセル情報の隣接セルへのキャッシュが可能である。これらの操作はノード間通信を含むため、通信が完了するまで後続の計算をブロックする必要がある。このため、しばしば通信ボトルネックを避けるための複雑な通信・計算のスケジューリング管理が必要となる。図は通信と各種計算の依存関係を示したものである。
本研究では、これらの通信を伴う命令と、通信結果を利用する後続計算の依存関係をセル単位で単純に表記する方法を提供した。加えて、上記依存関係に基づいて、システムは通信と計算のオーバーラップおよび各セルのデータを受信後直ちに後続の計算を実行できる機構を新たに導入した。
本研究内容は 2023 年に学会発表[2]にて発表済みである。



ソフトウェアの公開について：
研究（1），（2）に用いたソフトウェアおよび評価ベンチマーク群については，以下で公開中である。また（3）についても後日ソフトウェアを公開予定である。

- Handy Tools for Distributed Computing (HanDist): https://github.com/handist/
- 人工市場シミュレーション基盤 Plham (PlhamJ: https://github.com/plham/plhamJ)

５．主な発表論文等

〔雑誌論文〕　計6件（うち査読付論文　6件／うち国際共著　1件／うちオープンアクセス　4件）

１．著者名 Finnerty Patrick、Kamada Tomio、Ohta Chikara	４．巻 34
２．論文標題 A self‐adjusting task granularity mechanism for the Java lifeline‐based global load balancer library on many‐core clusters	５．発行年 2021年
３．雑誌名 Concurrency and Computation: Practice and Experience	６．最初と最後の頁 e6224
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cpe.6224	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 Finnerty Patrick、Kawanishi Yoshiki、Kamada Tomio、Ohta Chikara	４．巻 2022
２．論文標題 Supercharging the APGAS Programming Model with Relocatable Distributed Collections	５．発行年 2022年
３．雑誌名 Scientific Programming	６．最初と最後の頁 1～27
掲載論文のDOI（デジタルオブジェクト識別子） 10.1155/2022/5092422	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 －

１．著者名 Finnerty Patrick、Kamada Tomio、Ohta Chikara	４．巻 n/a
２．論文標題 Automatically balancing relocatable distributed collections	５．発行年 2023年
３．雑誌名 Concurrency and Computation: Practice and Experience	６．最初と最後の頁 e7717
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cpe.7717	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 Matsumoto Ryota、Kamada Tomio、Finnerty Patrick、Ohta Chikara	４．巻 11
２．論文標題 Topic-based distributed publish-process-subscribe system with metrics on geographic distance and permissible delay	５．発行年 2022年
３．雑誌名 IEICE Communications Express	６．最初と最後の頁 748～753
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/comex.2022COL0009	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 －

1．著者名 Tanaka Tomoya、Kamada Tomio、Ohta Chikara	4．巻 31
2．論文標題 Topic allocation method on edge servers for latency‐sensitive notification service	5．発行年 2021年
3．雑誌名 International Journal of Network Management	6．最初と最後の頁 e2173
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/nem.2173	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 該当する

1．著者名 Tanaka Tomoya、Kamada Tomio、Ohta Chikara	4．巻 9
2．論文標題 Distributed topic management in publish-process-subscribe systems on edge-servers for real-time notification service	5．発行年 2020年
3．雑誌名 IEICE Communications Express	6．最初と最後の頁 616～621
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/comex.2020COL0040	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4件（うち招待講演　0件／うち国際学会　3件）

1．発表者名 Patrick Finnerty, Tomio Kamada, and Chikara Ohta
2．発表標題 Integrating a global load balancer to an APGAS distributed collections library
3．学会等名 Proceedings of the Thirteenth International Workshop on Programming Models and Applications for Multicores and Manycores (PMAM '22)（国際学会）
4．発表年 2022年

1．発表者名 Yoshiki Kawanishi, Patrick Finnerty, Tomio Kamada, Chikara Ohta
2．発表標題 Distributed Cell Set : A Library for Space-Dependent Communication/Computation Overlap on Many Core Cluster
3．学会等名 The 14th International Workshop on Programming Models and Applications for Multicores and Manycores (PMAM `23)（国際学会）
4．発表年 2023年

１．発表者名
Patrick Finnerty, Yoshiki Kawanishi, Tomio Kamada, Chikara Ohta

２．発表標題
Experience in testing MPI+Java parallel and distributed programs with JUnit

３．学会等名
Summer United Workshops on Parallel, Distributed and Cooperative Processing (SWoPP2021), 2021

４．発表年
2021年

１．発表者名
Tomoya Tanaka, Tomio Kamada, Chikara Ohta

２．発表標題
Topic-based Allocation of Distributed Message Processors on Edge-Servers for Real-time Notification Service

３．学会等名
Proc. of The 21st Asia-Pacific Network Operations and Management Symposium (APNOMS 2020)（国際学会）

４．発表年
2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
Handist Collections （分散集合ライブラリ）
https://github.com/handist/collections
Handist Collections Benchmarks （ベンチマーク）
https://github.com/handist/collections-benchmarks
MPI Junit (MPI プログラムのテスト環境)
https://github.com/handist/mpi-junit
PlhamJ （人工市場シミュレーション基盤）
https://github.com/plham/plhamJ
Java GLB
https://github.com/handist/JavaGLB
Handist Collections
https://github.com/handist/collections

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	Ｆｉｎｎｅｒｔｙ　Ｐａｔｒｉｃｋ・Ｍａｒｔｉｎ (Finnerty Patrick Martin) (50957628)	神戸大学・システム情報学研究科・助教 (14501)	

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