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　 ５ 月 ２４ 日現在

機関番号：１２６０１
研究種目：若手研究
研究期間：2018～2022
課題番号：１８Ｋ１８０５９

研究課題名（和文）プロセス間負荷分散のための可変スレッド環境を提供する革新的なライブラリの開発

研究課題名（英文）Development of Dynamic Thread Mapping Library for Load-Balancing among Processes

研究代表者
河合　直聡（Kawai, Masatoshi）

東京大学・情報基盤センター・特任助教

研究者番号：８０７８０７９１

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：本件研究では、MPI/OpenMPのハイブリッドな並列環境を想定し、プロセス毎に異なるコア数を割り当てる(Dynamic Core Binding(DCB))ことで、プロセス毎の負荷の不均衡をスレッドレベルで均一化する手法およびDCB環境を提供するライブラリの開発を実施した。DCBを使用することで、負荷の不均衡があるアプリケーションの計算時間短縮または消費電力削減の効果が期待できる。実アプリケーションに適用してDCBの効果を検証した結果、スーパーコンピュータ上で最大で57%の計算時間短縮および77%の消費電力削減の効果を確認した。

研究成果の学術的意義や社会的意義
MPI/OpenMPのハイブリッドな並列が適用されたアプリケーションにおいて、均一な負荷分散の実現は困難な場合がある。そのようなアプリケーションに対してDCBを使用することで、負荷の均一化を実現し、計算時間短縮を実現する。また、使用するコア数を削減することで、消費電力削減も実現可能であることw確認している。DCBライブラリは簡易なインターフェイスで利用可能なため、スーパーコンピュータ上で計算されている様々なアプリケーションの計算時間短縮、消費電力削減につながると期待できる。

研究成果の概要（英文）：In this study, assuming a hybrid MPI/OpenMP parallel environment, we proposed a dynamic core binding(DCB) approach and developed a DCB library for reducing load-imbanace among the processes by changing the number of cores bound to each process. By using the DCB, we can be expected to reduce the computation time or energy consumption of applications with load imbalance. In the numerical evaluations with supercomputers, the computational time and energy consumption is reduced by the DCB library up to 57% and 77%, respectively.

研究分野：高性能計算

キーワード：Dynamic Core Binding　動的負荷分散　MPI/OpenMP ハイブリッド並列化　Power Aware

１．研究開始当初の背景

スーパーコンピュータを含む大規模な並列環境を効率的に利用するための並列化手法として、MPI+OpenMP によるハイブリッドな実装がある。MPI は並列単位（プロセス）毎に異なるメモリアドレスを管理しており、プロセス間のデータ共有にはプロセス間を記述する。このような特徴から、ノード内並列だけでなくノード間並列も実現可能である。対して OpenMP では、並列単位(スレッド)間で同一のアドレス空間を参照するため、ノード内並列化に限定されるが、MPIのような通信を記述することなく、スレッド間のデータ共有が可能となる。MPI と OpenMP は異なるメモリモデルを元に設計されており、それぞれ得意とする部分が異なるため、両方を併用した MPI+OpenMP のハイブリッド並列化（プロセス内でスレッド並列化を実施する手法）が広く利用されている。

MPI+OpenMP によるハイブリッドな並列化は実際に広く利用されており、様々なアプリケーションでその効果が実証されているが、しばしば負荷の不均衡が問題となる。並列化時に並列単位の負荷の不均衡がある場合、性能向上が阻害されるだけでなく、計算待ちが多く発生するために、無駄な電力が消費される問題も発生する。上述のように OpenMP はスレッド間で共有のアドレス空間を参照するために、均等な負荷分散は容易に実現できる。一方で、MPI ではデータ共有に通信を必要とするため、プロセス間のデータのやり取りに大きなオーバーヘッドが発生し、均等な負荷分散が困難となることが多い。

このような負荷の不均衡を解決する手法として、タスク並列化や、アプリケーションに特化した手法がある。タスク並列化は汎用的であり、容易に均衡な負荷分散を実現可能であるが、一般的にオーバーヘッドが大きく、性能向上が得にくい問題がある。また、アプリケーションに特化した手法は高い効果を得られるが、汎用性はなく、他のアプリケーションに適用しにくい問題がある。また、アプリケーション特化の手法は、MPI の特性から実装が複雑になるという問題もある。

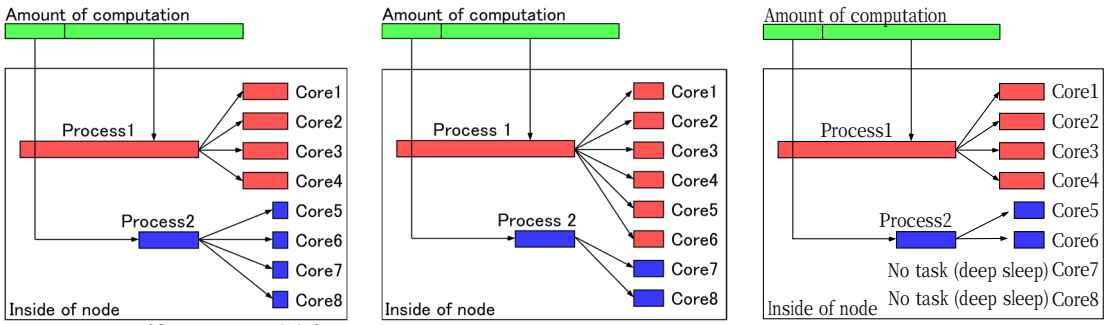
２．研究の目的

本研究の目的は、MPI+OpenMP で並列化された環境を想定した、均等な負荷分散を実現する手法および環境の実現である。具体的には、プロセスレベルでの負荷の不均衡をスレッドレベルで吸収する方法を検討する。

３．研究の方法

本研究では、プロセスレベルの負荷の不均衡を、プロセス毎に割り付けるコア数(スレッド数)を変えることで、スレッドレベルで均一化する手法(Dynamic Core Binding:DCB)[1]の提案および、DCB 環境を提供するライブラリの開発を行った。図 1 に DCB を使用していない場合と使用した場合のプロセスへのコア割り付けおよび、コアに割り付けられた演算量の違いについて示す。一般的な MPI+OpenMP 並列化では、プロセスに割り付けられるコア数は全て同じであり、図 1,a)に示すように、プロセス 1 にプロセス 2 の 3 の倍の演算量が割り付けられている場合、コアレベルでの演算量も 3 倍の差が出ている。

DCB ライブラリはプロセス毎の演算量に併せてプロセスに割り付けるコア数を変更することができる。図 1,b)に示すように、プロセス 1 とプロセス 2 に割り付けられるコア数を 3:1 になるように変更すると、コアレベルでの演算量が均一化され、全体の計算時間短縮が実現できる。また、近年のプロセッサは、コアに演算が割り付けられていない時の消費電力を大きく削減する機能が備わっているため、図 1,c)に示すように、プロセス 2 に割り当てるコア数を削減することで、全体の計算時間を変えることなく、消費電力の削減も可能となる。ここで、図 1,b)に示すような全てのコアを使用して計算時間を短縮する方法を DCB の RC モード、図 1,c)に示すような、演算量の最も多いプロセスに併せてそれ以外のプロセスに割り付けるコア数を減らし、消費電力を削減する方法を DCB の PA モードと表記する。



a) DCB を使用しない場合　　b) DCB:RC モード　　c) DCB:PA モード

図 1、DCB ライブラリの動作モードの違い

ここで、DCB で対応可能なのはノード内の負荷の不均衡だけであり、ノード間の負荷の不均

衡は別の方法で削減する必要がある。具体的には、プロセス毎の演算量を鑑みて、ノード毎に割り付けるプロセスの組み合わせを変更し、ノード間の負荷を均一にすればよい。ノードへの最適なプロセスの割り付け問題は、広義では組み合わせ最適化問題と呼ばれ、とりわけジョブスケジューリング問題に分類される。ここで、解くべき問題のノード数を 1,000、プロセス数を 4,000 と仮定すると、検討すべき組み合わせの数は数兆通りとなり、求解は困難である。そこで、本研究では、量子コンピューティングを採用した。量子コンピューティングは組み合わせ最適化問題に特化しており、近年様々な分野で利用が検証されている。ジョブスケジューリング問題でもその効果が実証されており[2]、本研究に適している。量子コンピューティングを使用して得た近似解を元に、ノードへのプロセスマッピングを変更し、DCB の効果を最大化する。

４．研究成果

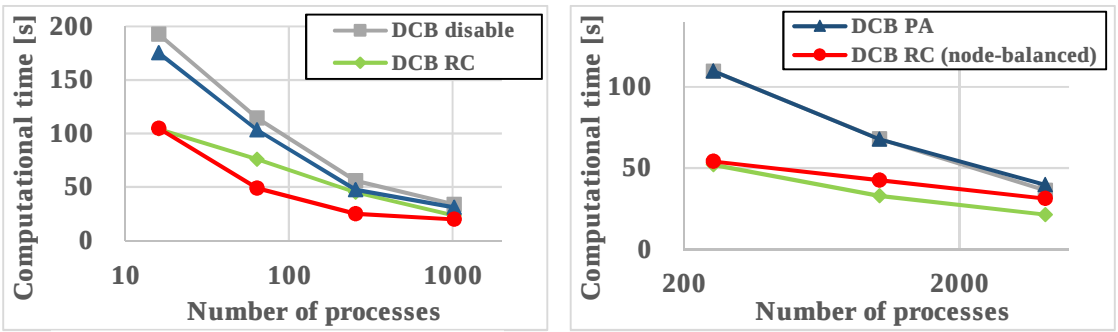
DCB ライブラリを実用的なアプリケーションに適用し、スーパーコンピュータ上で評価した結果、RC モードでは最大で 57%の計算時間短縮を、PA モードでは 77%の消費電力削減を実現した。

DCB の効果の評価は、実用的なアプリケーションの 1 つである、Lattice $\mathcal{H}$-matrix[3]に適用して実施した。Lattice $\mathcal{H}$-matrix は、密行列の近似表現であるオリジナルの$\mathcal{H}$-matrix を改良した手法である。密行列は様々なシミュレーションで頻出するが、大規模問題を扱う場合には多くの演算およびメモリリソースを消費するため、近似によって大きく計算時間、メモリ使用量を削減可能である。Lattice $\mathcal{H}$-matrix では MPI 並列化での通信時間削減に注力して改良が施されているが、演算量の不均衡が増大する問題が新たに発生しており、演算量不均衡の DCB による改善を実現する。

評価では東京大学情報基盤センターが運営しているスーパーコンピュータである OakBridge-CX:OBCX および Wisteria BDEC/01 Odyssey:WO を使用した。OBCX は Intel の Xeon CPU を採用した広く利用されている構成であり、WO は富岳と同様の富士通製 A64fx CPU を採用したシステムである。

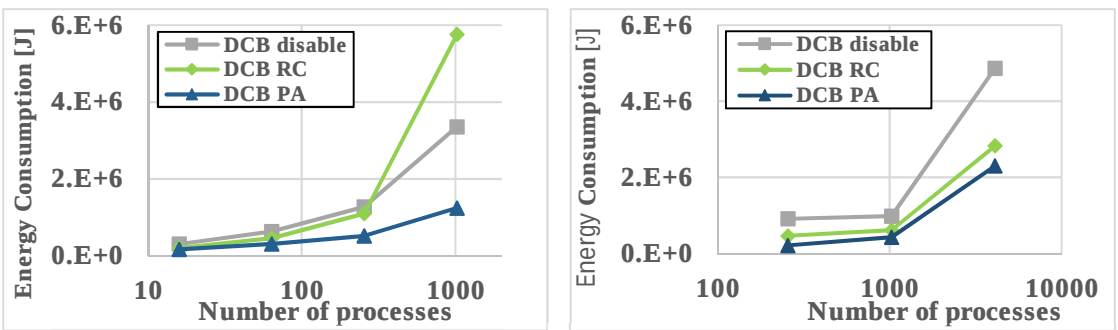
ノード間の負荷分散では、近似解を Fujitsu Digital Annealer:FDA を使用して取得し、近似解に基づいてノードに対するプロセスマッピングを変更した新たなコミュニケータを生成し、実現した。FDA 使用に際しては、解くべき問題を Quadratic Unconstrained Binary Optimization:QUBO に変換、pyQUBO にて実装した。

図 2、3 には OBCX、WO の各システムで DCB を使用した場合の計算時間および消費電力への影響を示している。DCB を使用しない場合（DCB disable）に対して、DCB の RC モードを使用した場合(DCB RC)には、いずれのシステムでも計算時間短縮の効果が確認できる。また、OBCX 上では、ノード間の負荷の均衡化を有効にした場合(DCB RC(node-balance))の計算時間がさらに短縮されている。一方で、WO 上では DCB RC(node-balance)の計算時間短縮の効果が確認できていない。これは組み合わせ最適化問題を設定する際に、スーパーコンピュータのネットワークの情報を含めておらず、通信パターンが複雑化し、通信時間が長くなった結果である。



a) Oakbridge-CX での評価結果　　b) Wisteria/BDEC-01 Odyssey での評価結果

図 2、計算時間に対する DCB の効果



a) Oakbridge-CX での評価結果　　b) Wisteria/BDEC-01 Odyssey での評価結果

図 3、消費電力に対する DCB の効果

次に、DCB の PA モードの効果を検討する。図 2 に示すように、DCB の PA モードを使用した場合(DCB PA)は、計算時間への影響はほとんどない。なお、OBCX で若干計算時間が短縮されているのは、使用するコア数の削減により、メモリバンド幅に余裕ができ、演算量の多いプロセスの計算時間が短縮されたためである。DCB の PA モードは想定通り、計算時間を変化させることなく、図 3 に示すように、いずれのシステムでも消費電力の削減を実現した。

[1] M. Kawai, A. Ida, T. Hanawa and K, Nakajima "**Dynamic core binding for load balancing of applications parallelized with MPI/OpenMP**" **International Conference on Computational Science(ICCS) 2023, Czech Republic, (In Press)**
[2] I. Attiya, A. E. Mohamed and X. Shengwu "**Job scheduling in cloud computing using a modified harris hawks optimization and simulated annealing algorithm**" **Computational intelligence and neuroscience 2020, (2020)**
[3] A. Ida "**Lattice H-matrices on distributed-memory systems**" **IEEE International Parallel and Distributed Processing Symposium (IPDPS). pp. 389–398 (2018)**

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　1件／うちオープンアクセス　1件）

１．著者名 Kawai Masatoshi、Ida Akihiro、Matsuba Hiroya、Nakajima Kengo、Bolten Matthias	４．巻 -
２．論文標題 Multiplicative Schwartz-Type Block Multi-Color Gauss-Seidel Smoother for Algebraic Multigrid Methods	５．発行年 2020年
３．雑誌名 Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region	６．最初と最後の頁 217-226
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3368474.3368481	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 Masatoshi Kawai, Akihiro Ida, Toshihiro Hanawa, Kengo Nakajima	４．巻 -
２．論文標題 Dynamic core binding for load balancing of applications parallelized with MPI/OpenMP	５．発行年 2023年
３．雑誌名 Computational Science; ICCS 2023	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 －

〔学会発表〕　計7件（うち招待講演　1件／うち国際学会　6件）

１．発表者名 Masatoshi Kawai
２．発表標題 Numerical Evaluation of Dynamic Core Binding Library with H-matrix Application
３．学会等名 Conference on advanced topics and auto tuning in high-performance scientific computing（招待講演）（国際学会）
４．発表年 2022年

１．発表者名 河合 直聡、伊田 明弘、中島 研吾
２．発表標題 不均一なコア割付による動的負荷分散手法の検討
３．学会等名 応用数理学会年会
４．発表年 2020年

１．発表者名
Masatosh Kawai

２．発表標題
Multiplicative Schwartz-type block multi-color GS smoother for AMG

３．学会等名
A French-German-Japanese workshop in Tokyo（国際学会）

４．発表年
2019年

１．発表者名
Masatosh Kawai

２．発表標題
The Effect of the Higher Precision on the IC Preconditioner

３．学会等名
Workshop on Large-scale Parallel Numerical Computing Technology（国際学会）

４．発表年
2020年

１．発表者名
M. Kawai, A. Ida and G. Wellein

２．発表標題
ppOpen-SOL: Robust ILU Preconditioner for Exascale

３．学会等名
HPC in Asia（国際学会）

４．発表年
2018年

１．発表者名
M. Kawai, A. Ida and K. Nakajima

２．発表標題
Higher precision for block ILU preconditioner

３．学会等名
CoSaS2018（国際学会）

４．発表年
2018年

１．発表者名
Masatoshi Kawai, Akihiro Ida, Toshihiro Hanawa, Kengo Nakajima

２．発表標題
Dynamic core binding for load balancing of applications parallelized with MPI/OpenMP

３．学会等名
International Conference on Computational Science 2023（国際学会）

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
本研究成果となるライブラリはbitbucketにて公開している。
https://naosou@bitbucket.org/naosou/dcb.git

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

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