

地球シミュレータ上でのXcalableMP言語の通信性能評価

上原均（海洋研究開発機構）、村井均（理化学研究所）、横川三津夫（神戸大学）

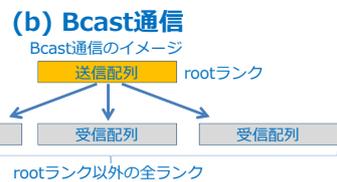
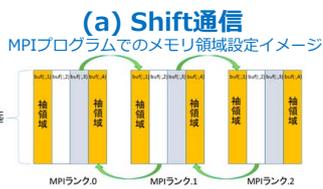
背景と目的

近年の計算機の著しい計算性能向上に従い、その並列度も急激に上昇しており、3~5年後に登場すると予測されるポストベタスケール級計算機では100万並列に達すると推定される。しかし一方で、そのような高並列な計算機を存分に活用するための高並列プログラムの開発は非常に難しい事が専門家から指摘されている。

そこで海洋研究開発機構と理化学研究所、神戸大学では、共同研究「ポストベタスケールシステム向けの並列計算モデルの開発と評価」を締結し、ベタスケール級および将来のポストベタスケール級のスーパーコンピュータにおいてアプリケーションプログラムを容易に高性能化するための並列計算モデル（プログラミング言語処理系）として、Partitioned Global Address Spaceモデルに基づいた並列プログラミング言語XcalableMP [1]（以下、XMP）について調査検討を行った。XMP言語は指示行ベースでFORTRAN言語あるいはC言語を拡張しており、生産性の高さが特に評価されている。上記の共同研究の一環として、本研究では、海洋研究開発機構が所有するスーパーコンピュータ「地球シミュレータ」上でXMP言語の利用可能性、特に高並列・大規模計算において性能面での影響が大きい通信性能について評価検討を行う。評価環境としてOmni XcalableMP Compiler Stable Version 1.1.0[2]を用いた。

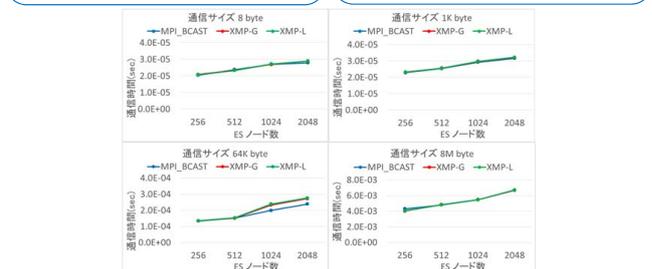
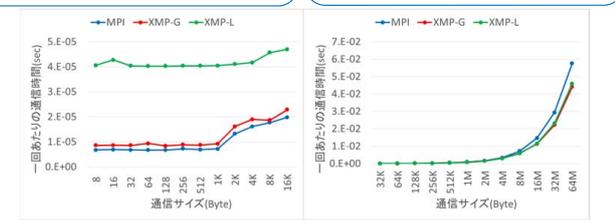
研究内容 (I) Shift通信、Bcast通信での評価

実アプリケーションで頻繁に用いられるShift通信とBroadcast(以下、Bcast)通信について、従来使われてきたMessage Passing Interface (MPI)での実装と、XMP言語での実装を比較する。XMP言語では各ノードで共有して利用するデータを各ノードに分散配置するGlobal View方式と、各ノードが持つローカルデータに対して通信を行うLocal View方式がある。本研究では両方での実装を行って比較検討する。Bcast通信では地球シミュレータの特徴である拡張クラスタ（2048ノード）での高並列時の性能評価も行う。なお以下の計測は各ノード1プロセスで実行しており、計測区間前後ではバリア同期を行っている。性能値は100回試行での平均値である。



- XMP Global Viewへの書換え
 - データ分割は指示行で指定
 - MPI_Send等をreflect構文で書換え
 - 通信区間をサブルーチン化（※プログラムの仕様上、reflectされるデータ長が動的に変わるため）
- XMP Local Viewへの書換え
 - MPI_Send等をCAFの片側通信に書換え
 - 例：`buf(1:len,4)=buf(1:len,2)[rank+1]`
 - ※rank+1のプロセスのbuf(1:len,2)を自分のbuf(1:len,4)にコピー

- XMP Global Viewへの書換え
 - MPI_Bcast関数はBCAST指示文で書換え
 - XMPのBCAST指示文は指定した配列の全要素を送信するため、動的に通信バッファを確保する方式に変更
- XMP Local Viewへの書換え
 - MPI_Bcast関数はCo-array Fortran (CAF)のCO_BROADCAST関数に書換え
 - CO_BROADCAST関数は指定した配列の全要素を送信するため、動的に通信バッファを確保する方式に変更

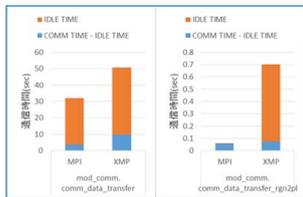


短メッセージ時：MPIが良好な性能
長メッセージ時：XMP Global View記述が良好な性能
※XMPの内部実装でMPI_Send_Init関数等を利用

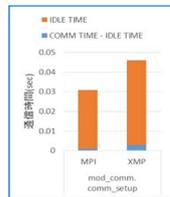
Bcast通信ではMPIとXMPは高並列でもほぼ互角の性能

研究内容 (II) NICAM-DC-mini XMP版での評価

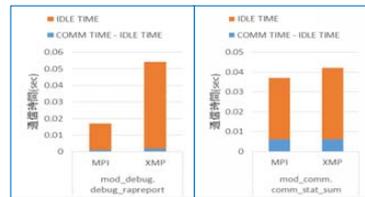
理化学研究所計算科学研究機構が公開しているFiber[3]に含まれるNICAM-DC-miniのMPI版とXMP言語版[4]について、地球シミュレータでの実行性能、特に通信性能に着目して比較評価する。評価にはjablonowski (GLevel05, RLevel00, 40層, 10並列)を用いた。NICAM-DC-miniの通信処理は1) 隣接間1対1通信、2) ブロードキャスト通信、3) リダクション演算を伴う集団通信、に大別される。XMP言語版はCo-arrayによるLocal View方式で記述されている。



隣接間通信部分の性能評価例
XMP実装でのオーバーヘッドの問題



ブロードキャスト通信部分の性能評価



リダクション演算を伴う集団通信部分の性能評価例
XMP実装でのオーバーヘッドおよび演算負荷の不均衡

まとめと今後の課題

- 長メッセージのShift通信ではMPI版よりXMP (Global View) 版が**性能的に優位**
- Bcast通信ではMPI版とXMP (Global View) 版は**性能的にほぼ互角**
- XMP (Local View) 版は**性能的にはやや不利だがCAF記法の可読性は高い**
- NICAM-DC-mini XMP版は通信時間が増加傾向
⇒ **オーバーヘッドについてXMP言語の実装に改善の余地あり**
- 今後の課題：より本格的なコードや周辺プログラム（例：可視化処理）等への適用

謝辞

本研究の一部は、文部科学省フラッグシップ2020プロジェクト（ポスト「京」の開発）「ポスト「京」で重点的に取り組むべき社会的・科学的課題」における重点課題④「観測ヒューマンを応用した気象と地球環境予測の高度化」として実施されました。また、本研究の実施ではNECソリューションイノベータ株式会社 山口健太氏と関係各位の協力を得ました。ここに記して感謝の意を表します。

参考文献

- XcalableMP, <http://www.xcalablemp.org/>
- Omni Compiler, <http://omni-compiler.org/>
- Fiber, <http://fiber-miniapp.github.io/>
- PGAS言語XcalableMPによるFiberミニアプリ集の実装と評価、村井均、2016年ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2016)