

MODYLAS を用いた FX100 のアシスタントコアの性能評価

中島 大地[†] 安藤 嘉倫[‡] 坂下 達哉^{†3} 片桐 孝洋^{†4} 荻野 正雄^{†4} 永井 亨^{†4}

名古屋大学工学部電気電子・情報工学科[†] 名古屋大学 工学研究科附属計算科学連携教育研究センター[‡]
名古屋大学 工学研究科^{†3} 名古屋大学 情報基盤センター^{†4}

1. はじめに

MODYLAS (MOlecular DYnamics software for LARge System) [1]は、巨大系に適した高並列汎用分子動力学 (MD) シミュレーションソフトウェアであり、現在は物理学、化学、生物物理学といった分野において広く利用されている。

MODYLAS は様々な MD 計算手法と、階層的並列化手法を実装している。数万プロセスに対応した並列化 MD 計算を MPI/OpenMP ハイブリッド並列や新たなデータ構造を用いることで実現している [1]。しかし現状では、MPI の通信オーバーヘッドなどに改善の余地がある。

本研究では、MPI 通信部分の性能改善を検討するにあたり、名古屋大学に設置したスーパーコンピュータ FX100 システム (以降、FX100) を用いて性能評価を行う。FX100 の Tofu 通信網に備わっている 4 レーン環境と 2 レーン環境の 2 つについて、実プログラムでの MPI 通信性能の違いを評価する。さらに FX100 から実装されているアシスタントコアを明示的に利用した際に、非同期通信処理時における通信隠蔽による台数効果がどの程度生じるかを検証する。以上の 2 機能について現状の MODYLAS は十分に最適化されておらず、性能評価に基づく最適化により MPI 並列性能の向上が期待できる。さらに多様な性能に関するパラメータを最適化し、MODYLAS のストロンングスケール性能を向上させることを本研究の目的とする。

2. 通信ネットワーク Tofu2 とアシスタントコア

FX100 の通信ネットワークである Tofu2 によるノード間通信は、4 レーンを有する。バンド幅はリンク当たり 12.5GB/s である。名古屋大学情報基盤センターでは運用時の安定性を考慮し 2 レーン運用がされている。2017 年度、試験運用として 4 レーンの提供がされている。そこで本研究ではネットワーク性能の違いによる通信性能を評価するため、2 レーンと 4 レーン を利用する。

2 レーンと 4 レーンをそれぞれ用いて、MODYLAS はリンク幅性能の違いによるノード間の通信時間の短縮の恩恵を受けるのか、またそれらの影響について性能評価を行う。

アシスタントコア [2] [3] は、FX100 に実装されている、通信や OS の処理を専用に扱うコアである。1 ノード上には 2 つのアシスタントコアが実装されている。本研究においては、MODYLAS を利用し、非同期通信処理における通信時間隠蔽機能がより有効となるかを検証する

3. 性能評価環境

性能評価は名古屋大学情報基盤センターに設置された FX100 を用いて行う。使用される入力データは、MODYLAS1.04 に含まれる water のサンプルと、water を最小化して作成したデータを利用した。また使用ノード数は、MODYLAS の仕様上の制約から 2 のべき乗であること、および試験運用中の 4 レーン環境の最大ノードは 12 ノードであることから、8 ノードとした。またノード次元形状は 2x2x2 とした。

実験は、(1) 通信レーンを 2 レーン / 4 レーンを用いた場合、(2) 非同期通信時にアシスタントコアを用いた促進機能を用いた場合 / 用いない場合、の 2 種類の観点から行う。それぞれ平均値、最大値、および最小値を評価した。実験環境設定について表 1 にまとめる。

表 1 異なる通信レーンの比較を行う環境設定

ノード割り当て	離散割り当て
通信方式	同期通信
観測ステップ数	100000 ステップ
観測回数	10 回
計算方法	エネルギー最小化

まず、通信レーン数を変化させたときの通信時間について評価する。図 1 に表 1 に基づいた結果を示す。図 1 から、2 レーンと比較し、4 レーンは実行時間を短縮することができるといえる。実行時間の変動幅 (最大値と最小値の差) についても、4 レーンでは 2 レーンに対して、より小さくなる。これは、通信レーン数の増加により通信の輻輳が起きず、結果として高速化されたことが理由と考えられるが、詳しい解析が必要である。

Performance Evaluation of Assistant Cores on the FX100 with Modylas

[†] Daichi Nakajima, Electronic Engineering and Information Engineering, school of Engineering Nagoya University

[‡] Yoshimichi Andoh, Center for Computational Science, Graduate School of Engineering Nagoya University

^{†3} Tatsuya Sakashita, Graduate School of Engineering Nagoya University

^{†4} Takahiro Katagiri, Toru Nagai, Masao Ogino, Information Technology Center, Nagoya University



図 1 レーン数の違いによる実行時間比較

またノード割り当てについては、メッシュ割り当て（物理的に連続するノード割り当て）の実行と比較して、離散割り当ての方が4レーン実行の効果があった。この理由は、メッシュ割り当てが隣接するノードを確保する一方で、離散割り当ては隣接するノード以外を用いてノードの確保をすることがあるため、離散割り当てのほうがメッシュ割り当てよりも通信数が増加することがある。そのため、レーン数の多い4レーンが有効となる可能性があるからと考えている。

次に非同期通信での明示的なアシスタントコア利用時の性能について評価する。表2にアシスタントコア利用時の環境設定を記載する。

表 2 アシスタントコア利用時の環境設定

ノード割り当て	メッシュ
通信方式	非同期通信
観測ステップ数	50000 ステップ
観測回数	5 回
計算方法	NVT アンサンブル計算

結果を図2に示す。図2から、アシスタントコアを用いた場合、実行時間が増加するという結果となった。特に1ノード内のプロセス数が増加すると性能劣化が顕著になった。この理由は、1ノード間にアシスタントコアが2コアしか存在しないため、2コアのアシスタントコアを複数のプロセスで共有したことにより性能が劣化するから[2]と考えられる。

またこの影響がないはずの1ノード内に2プロセスしか存在しない場合においても、アシスタントコアの積極的な利用で高速化されることはなかった。この理由として、アシスタントコアによる非同期通信による通信隠蔽の促進の効果よりも、ハイブリッドMPI/OpenMP実行においてスレッド数を増やすことによる演算効率の低下のほうが大きくなったことが考えられるが、性能プロファイラによる詳細解析が必要である。

4. まとめ

FX100で実装されている専用ネットワークTofu2によるノード間通信では、レーン数の増加

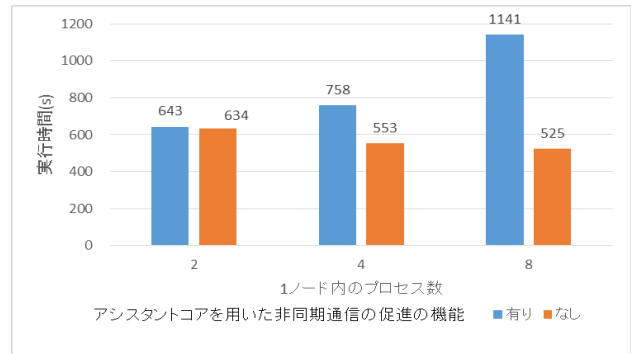


図 2 アシスタントコアによる実行時間比較

による効果は、ノード割り当てについて離散割り当てを用いるときに効果があった。一般的に、ノードを連続するように確保するメッシュ割り当ては、活用できる計算資源の空きがないときには、ジョブ開始までに時間がかかるため、ターンアラウンド時間の観点では離散割り当てがよい場合がある。このように、離散割り当てを利用せざるを得ない状況下では、離散割り当てを用いて4レーン実行を行うことで、結果として2レーン実行よりもターンアラウンド時間を短縮できる可能性がある。

一方、本評価においてはアシスタントコア利用の有能性が認められなかった。この理由の1つは、現在のMODYLASの非同期通信が、通信と演算のオーバーラップを意識して実装されていないためである。さらに今回の入力規模および並列実行規模において、1ノード内に2プロセスを利用する実行よりも、4プロセスや8プロセスを利用する実行形態のほうが、実行時間が短くなった。より大きな入力を用意するなどして、ノード内2プロセスを想定したFX100のアシスタントコアの設計方針と合致した系で再測定する必要がある。以上の問題を考慮してMODYLASのコード最適化を行い、プロセス数を考慮したアシスタントコアの性能評価を行うことが今後の課題である。

謝辞

本研究の一部は、学際大規模情報基盤共同利用・共同研究拠点の支援(課題番号: jh170024-NAH)、JSPS 科研費 16H02823 の助成による。

参考文献

- [1] 2013 MODYLAS Group, MODYLAS 1.0.2 Reference Manual (2014/10/4), MODYLAS 1.0.2 Tutorial (2014/10/4), 2014. <http://www.modylas.org/node/1>
- [2] 富士通株式会社, FUJITSU Software Technical Computing Suite V2.0, MPI 使用手引書 (PRIMEHPC FX100 用), 2017.
- [3] 富士通株式会, FUJITSU Supercomputer PRIMEHPC FX100 先進のソフトウェア, 2014. <http://www.fujitsu.com/downloads/JP/archive/ingjp/jhpc/primehpc/primehpc-fx100-soft-ja.pdf>