jh150015-NA11

分子動力学計算ソフトウェア MODYLAS の メニーコアアーキテクチャ対応並列化に関する研究

安藤 嘉倫（名古屋大学 工学研究科 計算科学連携教育研究センター）

概要　本研究では、汎用分子動力学計算ソフトウェア MODYLAS について、次世代のメニーコアアーキテクチャ（FX100 および Xeon Phi）の性能を発揮させるための並列化チューニングを行った。本年度は特にスレッド並列計算について、まず問題点抽出のためのコード性能評価を行った。その上で MODYLAS を用いた分子動力学計算における２つのホットスポット（粒子対での相互作用計算 p2p および高速多重極展開法での多極子展開係数から局所展開係数への変換 M2L）でのスレッド間ロードンバランスの解消を目的として、スレッド並列性能向上のための並列アルゴリズム開発およびコーディングを行い、顕著な性能向上を実現した。

1.　共同研究に関する情報

(1) 共同研究を実施した拠点名

名古屋大学情報基盤センター
東京大学情報基盤センター

(2) 共同研究分野

口　超大規模数値計算系応用分野

(3) 研究者の役割分担

- 全体統括 安藤
- MPI 並列化性能統括 荻野
- 分子動力学計算のアルゴリズム開発
  安藤、吉井、藤本、遠藤、篠田、岡崎
- 並列化コーディング（MD 計算全般）
  安藤，藤本，遠藤
- プログラム性能評価 大島、片桐
- Xeon Phi 向け並列化コーディング 大島
- 自動性能チューニング技術提供 片桐
- 並列化コーディングおよび並列アルゴリズム開発 小村、鈴木

2.　研究の目的と意義

研究の目的　分子動力学 (MD) 計算は，化学，物理，生物，およびウィルス学といった様々な学問分野において実験とならぶ解析ツールとして広く普及している．加えて工業分野においても分子の特性を活かしたナノ機能性材料や高分子材料 (図 1) を設計する際に MD 計算により得られる知見が不可欠になりつつある．しかしながら，長距離原子間相互作用を含めた実用の研究において MD 計算であつかえる原子数および計算時間は，「京」コンピュータといった最新鋭のスーパーコンピュータを用いたとしても 1 千万原子系 (空間サイズとして 50 ナノメートル立方程度) に対する数 100 ナノ秒の計算が限界である[1]．より大規模かつ長時間な MD 計算を行うことで，学問上のブレークスルーだけでなく，より高精度な材料設計が可能になると期待される．

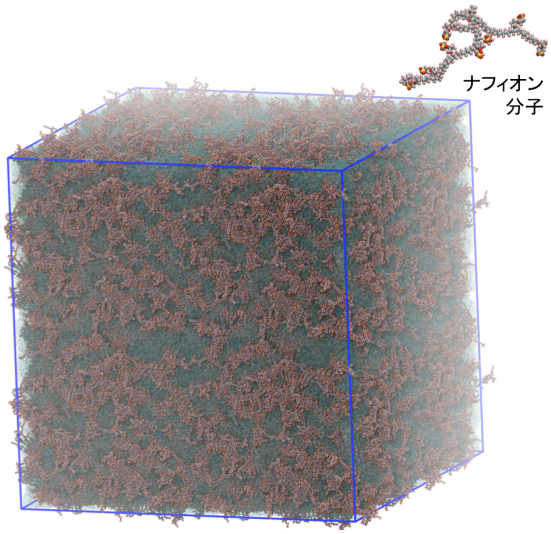

図 1　MODYLAS を用いた計算対象の例。燃料電池に用いられるナフィオン高分子分離膜 (全原子モデル)。

本研究では「京」コンピュータおよび FX10 上で稼働実績のある汎用分子動力学ソフトウェア MODYLAS[2,3]に対して，次世代のメニーコア計算機，さらには将来のエクサスケールマシンの性能を発揮させるための並列チューニングを行った．必要に応じてハードウェアに適合したアルゴリズム開発についても行い，将来的にエクサスケールマシンと MODYLAS によって長距離原子間相互作用を含めた数億から 10 億原子系での実用的な MD 計算を可能にすることを目標とする．

研究の意義　分子動力学ソフトウェア MODYLAS は「京」コンピュータの全ノード規模での並列化計算に対応した MPI/OpenMP ハイブリッド並列化チューニングがすでに施されており，次の目標は数 10 PFLOS (ポスト T2K) ないしエクサスケールマシンでの高効率な並列化計算への対応にある．一方スーパーコンピュータの進化はノード数の増加が頭打ちになり，ノード当たりのコア数および SIMD ベクトル長を増加させることでシステム全体の演算性能を向上させる方向にある．これら次世代のマシンでは，コア数および SIMD 長の増加によって現行のプログラムの各所でスレッド並列性および SIMD 並列性が確保できなくなる。本研究でのスレッド並列性確保のためのチューニング，計算負荷の均等化によるスレッド並列化効率の向上，および階層キャッシュの最適化についての研究は，エクサスケールマシン上で MODYLAS による高効率なスレッド並列化 MD 計算を実現するために不可欠である。

本研究では、FX100 をベースに 32 コアに対応したスレッド並列化コーディングを実現した上で，さらなるコア数増加に対応できるよう Xeon Phi をベースにした 50 コア以上のスレッド並列化を見越した研究を行った。研究成果は顕著であり，今後研究をより深化させることでポスト T2K およびエクサスケールマシンでのメニーコア並列環境へスムーズに対応できると期待される．さらに大規模なハイブリッド並列においてプロセス数とスレッド数の最適化は様々なチューニングパラメータの関与する複雑なプロセスであり，本研究成果をベースとした今後の研究において，このコストを自動性能チューニング技術によって削減することを図る．近い将来、次世代スーパーコンピュータを高効率に動作させ，数億～10 億原子系での MD 計算を数 100 ナノ秒～数 μ 秒のオーダーで行うことにより，上記学問分野におけるより現実的な MD シミュレーションにもとづく新規な科学的発見，および工業分野における実用的な MD シミュレーションによる材料設計に貢献する．

3. 当拠点公募型共同研究として実施した意義

MD 計算ソフトウェアがハードウェアの日進月歩の進化に対応するためには，最新のアーキテクチャの基本性能および対応コーディングに詳しいコンピュータ・サイエンスの研究者と，MD シミュレーションを用い実際の研究を実施している者とが，互いに協力した学際領域分野の研究が不可欠である．本研究では両分野の専門家が共同研究者として参画している．本課題に参加するコンピュータ・サイエンスの研究者はハイパフォーマンスコンピューティング(HPC)全般を専門とし，特に自動性能チューニング(AT)に詳しい研究者（片桐)、最新のメニーコア技術に詳しい研究者(大島),スレッド並列(小村,鈴木)および MPI 並列化技術に詳しい研究者(荻野）から構成される．複合・階層的な最新のメニーコア型の計算機システムを使いこなす上で不可欠な人員構成である。一方，MD シミュレーションの研究者は，MD 計算ソフトウェア MODYLAS の基本設計に最初から関与しかつ MD 計算の基本原理に詳しい研究者(安藤,吉井,岡崎,篠田)および現行の MODYLAS 並列化の内容に詳しい研究者(安藤,藤本,遠藤)から構成される．両者の協業に

より，プログラミングレベルでの並列性能チューニングだけでなく，新規並列アルゴリズム開発・実装による性能向上を迅速に達成することができた.

4. 前年度までに得られた研究成果の概要

本年度よりの開始課題のため該当せず。

5. 今年度の研究成果の詳細

当初 4 月 1 日より利用可能予定であった名大情報基盤センターの FX100 の運用が 9 月にずれ込んだため、スレッド並列の事前評価をまず東大情報基盤センターの FX10 上，および Xeon Phi 上で行った。

MODYLAS では長距離静電相互作用を高速多重極展開法 (FMM，図 2)を使って計算している。MD 計算においてのホットスポットは以下の二箇所，

- Lennard-Jones 相互作用および静電相互作用の粒子対計算 p2p [energy_direct]
- 多極子から局所展開係数への変換 M2L [energy_fmm_2 および energy_fmm_3]

である。ここで[...] 内はサブルーチン名。この二箇所を本課題でのスレッド並列最適化の対象とした。

5.1 FX10 および FX100 での最適化

まず energy_direct について，基本的性能を調べるため FX10 での 16 スレッド実行時の演算性能およびスレッド間インバランス性能の測定を行った。そのデータを元に問題点を抽出しスレッド並列性能改善に取り組んだ。

図 3 にあるように，energy_direct のオリジナルコードでは 16 スレッド実行時にスレッド間に最大 30%のインバランスが生じていた。その原因として図 5 左にあるように，オリジナルコードでは対相互作用の自原子 (iatom) に関するループが iatom の属する小領域 (サブセル) 内の原子数 (平均 40 程度) しかなく，16 スレッド実行時に粒度不足に陥っていた。参考のため図 6 上にはオリジナルの energy_direct 疑似コードを示す。この粒度不足を解消するため 2 つの観点からスレッド並列の粒度向上および負荷均等化を試みた.

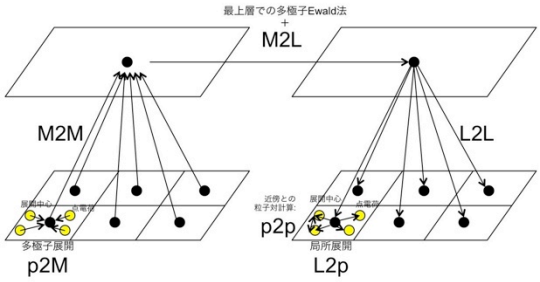

図 2 高速多重極展開法 (FMM)の主要演算．ホットスポットは近傍粒子との粒子対計算 p2p [energy_direct] および多極子から局所展開係数への変換 M2L [energy_fmm_2 および_3]．演算はほか点電荷から多極子展開係数への変換 p2M [calc_fmm]，多極子のマージ M2M [merge_fmm]，局所展開中心の移動 L2L [energy_fmm_2 および_3]，および局所展開係数をもちいた点電荷上の電場の計算 L2p [energy_fmm の一部] から構成される。

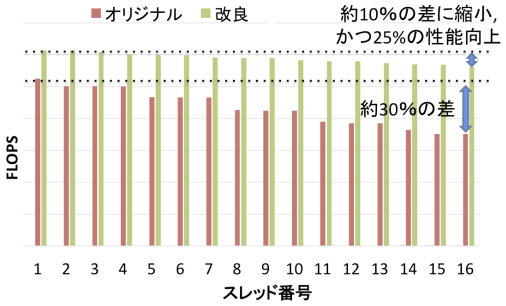

図 3　FX10 での energy_direct 16 スレッド実行時のスレッド番号別演算性能。改良その 1 により 30%のインバランスが 10%に削減された。

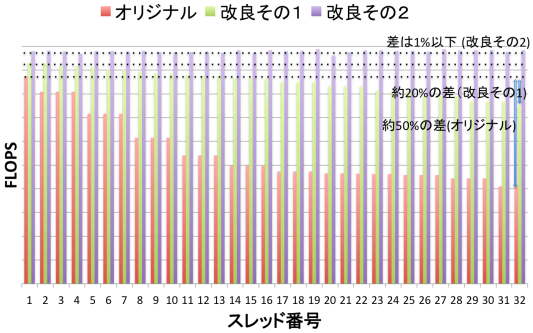

図 4　FX100 での energy_direct 32 スレッド実行時のスレッド番号別演算性能。改良その 2 によりインバランスは 1%以下に削減された.

一つ目の観点は iatom についてのループの伸張である．図 5 右にその概念図を，図 6 下に具体的なコードを示す．該当 do ループにて参照する小領域を z 軸方向に複数のサブセルにまたがるよう広げることで iatom に関するループ長を伸張した。コード改良の結果，図 3 にあるように 16 個のスレッド間でのロードインバランスが 10% に縮小され，かつループ長伸張の効果によりスレッド単体性能についても 25%程度向上した。この方法では相手 jatom 粒子についての参照範囲も拡大するためキャッシュミス率の増大が心配されたものの，現状深刻な問題とはなっていない。

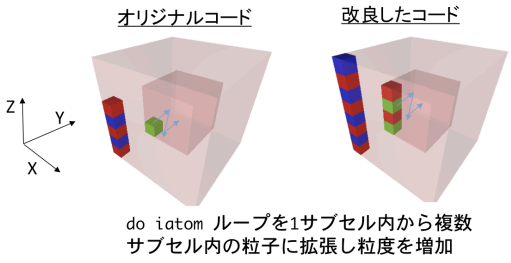

図 5　energy_direct コード改良の概念図．中心の緑および赤い立方体が iatom の所属セル，端の赤および青い立方体が jatom の所属セル．

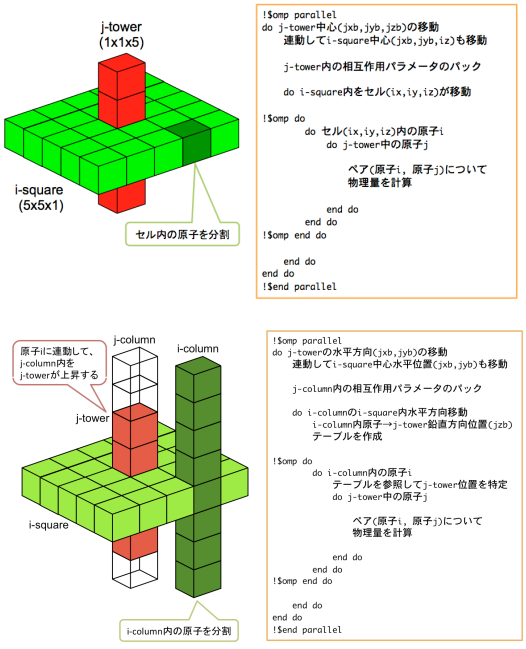

図 6　energy_direct の疑似コード。オリジナルコード(上)，および改良コードその 1 (下)

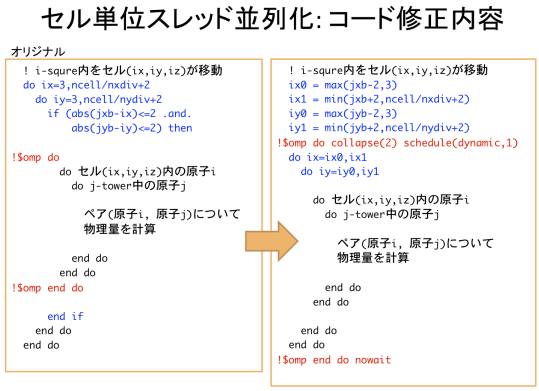

図 7　energy_direct の疑似コード。改良コードその 2。

二つ目の観点からの改良では，スレッド並列対象箇所を iatom のループではなく iatom の属するサブセルに関するループに移動した．その具体的なコードを図 7 に示す．この方法では，分割対象ループに nowait を指定しているためスレッド数の上限は [MPI プロセス内のサブセル数 N_{sub}]×25 になる．$N_{sub}=2^3$ のとき 200，4^3 のとき 1600 であり，メニーコアに対して十分な並列度が確保される。図 4 にあるように，32 コア実行時 ($N_{sub}=4^3$) でのスレッド間ロードインバランスが 1%以下と劇的に縮小され，かつスレッドあたりの計算粒度が約スレッド数倍になった効果などによりスレッド単体性能についても向上した。その結果 $N_{sub}=4^3$ のとき 1.5 倍 (16core) および 1.7 倍 (32core) の高速化が実現された．我々はさらに性能を向上すべく i-square 単位，i-column 単位のスレッド並列化についても検討しており，問題規模ごとのこれら方法の最適解を探っている最中である．

もう一方のホットスポット energy_fmm_2 についてはプロセス数に応じて M2L 演算を FMM 階層ごと異なる相互作用計算対象セル数 (すなわち，do ループ長さ) で行うという特徴がある。しかしながら、オリジナルコードではスレッド並列のチャンクサイズを階層によらず一定かつ 8 スレッド実行に最適化した値としていた。そのため並列対象 do ループ

粒度の大きい下層部についてはインバランスが小さいが, 粒度の小さい上層部ではインバランスが大きくなる傾向があった。階層ごとチャンクサイズの最適化を行うことでこのインバランスを大幅に解消することができた (5.2 節で後述)。

ほか、FX10 での基本プロファイル測定データより明らかになった問題点として、スレッド間のバリア同期の回数の多さとそれに伴うバリア同期に要する時間の割合が大きい点があった。例えば図 8 にあるように, 主要ルーチンの p2p 演算 energy_direct で 12%, L2p 演算 energy_fmm で 56%, p2L 演算 calc_fmm で 55%のバリア同期があり, プログラム全体でも 12%と非常に大きかった。バリア同期が多い原因として, これらサブルーチンにおけるリダクション指示句を使った変数 (例えばポテンシャルエネルギー) の足し込みがあると考え, 対象変数をスレッド数の次元を持つ配列として変数の足し込みをループ外で行うよう書き換えた。さらに, 依存性を吟味した上で該当 do ループ末に nowait 指示句を入れる, 特に energy_direct についてはスレッド並列対象の do ループ位置を変える (図 7), などの対処を施した。その結果、図 9 にあるように energy_direct および energy_fmm におけるバリア同期を大幅に減少させることができた。しかしながらプログラム全体のバリア同期の割合は依然 9%と高く, さらなる性能向上にはここで行ったと同様の改良をプログラムの該当箇所に広く適用する必要がある。

また詳細プロファイル測定データから, 主要ルーチン以外で L1 キャッシュの dm ミスヒット率が顕著なサブルーチンが残されていることがわかった。これらルーチンについてデータの事前ソートによる不連続性の解消, およびデータ量のインバランスの調整による改善を計画している。さらに MIPS 値測定からサブルーチン energy_fmm, calc_fmm などで整数演算が目立っており, そこで使用されているユーザー定義関数 algndr (ルジャンドル倍多項式) での整数演算の効率化, および if 文の削減についても検討している。

基本プロファイル結果(pyp222)

```
Application - procedures
        Cost      %        Operation (S)     Barrier      %   Start     End
-------------------------------------------------------------------------------
       27709 100.0000       2772.3811        3360     12.1260    --      --    Application
-------------------------------------------------------------------------------
       10847  39.1461       1085.2798        1325     12.2154    98     334    energy_direct._OMP_1_
        7103  25.6343        710.6797         151      2.1259  1345    1445    energy_fmm_2._OMP_1_
        1914   6.9075        191.5023        1072     56.0084  1142    1251    energy_fmm._OMP_1_
        1347   4.8612        134.7720           0      0.0000    --      --    __g_dscn
        1121   4.0456        112.1599           0      0.0000  2922    2958    algndr_
         783   2.8258         78.3419           0      0.0000    --      --    __g_dexp
         700   2.5263         70.0374         388     55.4286   241     325    calc_fmm._OMP_1_
         417   1.5049         41.7223           0      0.0000    --      --    __g_cdexp
         271   0.9780         27.1145          14      5.1661   278     445    pshake_roll_main_
         234   0.8445         23.4125          81     34.6154  1449    1488    energy_fmm_2._OMP_2_
```

図 8　FX10 上での基本プロファイル測定結果。インプット pyp222 はタンパク質, イオン, および水を含む約 15 万原子系。8 プロセス 16 スレッド実行。赤色が性能上改善の余地のある箇所。

```
Application - procedures
        Cost      %  Operation (S)     Barrier      %   Start     End
-------------------------------------------------------------------------------
       15188 100.0000       1539.3612        1435      9.4482    --      --    Application
-------------------------------------------------------------------------------
        5472  36.0284        554.5701           0      0.0000  1357    1457    energy_fmm_2._OMP_1_
        3787  24.9342        383.7962          76      2.0069   417     643    energy_direct._OMP_1_
        1095   7.2096        110.9755           0      0.0000    --      --    __g_dscn
         743   4.8920         75.3178           0      0.0000  2939    2975    algndr_
         729   4.7998         73.8971           0      0.0000  1146    1258    energy_fmm._OMP_1_
         691   4.5496         70.0330         602     87.1201   241     324    calc_fmm._OMP_1_
         476   3.1341         48.2434           0      0.0000    --      --    __g_dexp
         463   3.0485         46.9325           0      0.0000    --      --    __g_cdexp
         184   1.2115         18.6623         137     74.4565  1540    1665    remove_void123lj._OMP_1_
         165   1.0864         16.7161         135     81.8182   564     695    m2m._OMP_1_
```

図 9　FX100 上での基本プロファイル測定結果（改良コードその 2)。インプットは図 10 と同じ。8 プロセス 16 スレッド実行 ($N_{sub}=4^3$)。青色がバリア同期削減に成功した箇所。

5.2 Xeon Phi での最適化

図 10 には energy_direct, energy_fmm_2, および energy_fmm_3 へ上記改良を加えた上での Xeon Phi 上 60, 120, 180 および 240 スレッド実行での性能測定結果を示す。さらに図 11 には例として 240 スレッド実行時のスレッド番号別経過時間を示す.

energy_direct オリジナルコードについてスレッド数の増加にともない経過時間が増加した(図 10). これは図 11 上にあるように do ループ回転数の枯渇によるスレッド間負荷インバランスが発生したためである。今回の改良その 1 を施すことによって 120 スレッドに至

るまでの高速化が達成されたが, 180 スレッド以上では飽和した (図 10)。また FX100 上では改良その 2 が性能的に優位であったが, Xeon Phi 上では反対の傾向となった。予想に反する結果のためその原因については現在調査を進めている。

一方 energy_fmm_2 および_3 について, オリジナルコードでは最下層および二階層目については均一なスレッド並列性が確保される一方, より上層においてはチャンクサイズの不備によりインバランスが生じていた(データ未掲載)。今回チャンクサイズの最適化を行った結果, 図 11 下にあるように 3 階層目以降においても均一な並列性を確保できるようになった。ただし 240 スレッドすべてを活用できておらず, チャンクサイズの計算式には改善点がのこる。今後 Xeon Phi のコア数がさらに増えた場合, energy_fmm_2 および_3 ともに並列性を保つ上で該当 do ループ長のさらなる伸張は不可欠である。しかしながら 1 階層ごと M2L を行う現在の実装では, 相互作用計算対象セルのループ長は数千程度に限られている。メニーコアをより有効に活用する改良の候補として, ループ融合により階層を跨いだループに書き換えるなどを考えている。

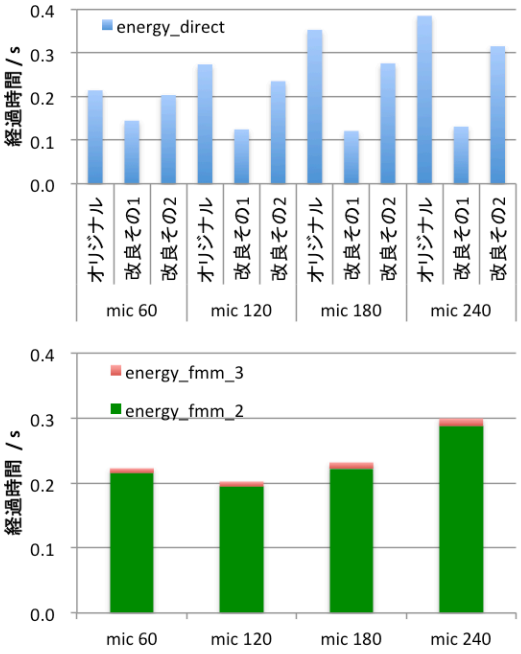

図 10 Xeon Phi での energy_direct (p2p), energy_fmm_2, energy_fmm_3 (M2L+L2L) の測定値. energy_fmm については階層ごとの合計値. スレッド数 60, 120, 180 および 240.

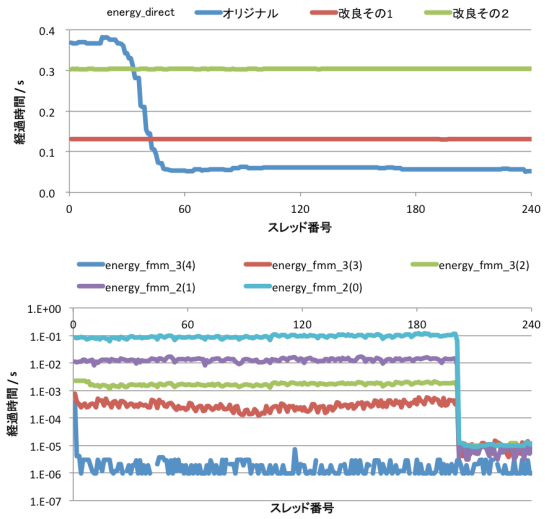

図 11 Xeon Phi での energy_direct (上), energy_fmm_2 および energy_fmm_3 (下) 測定値。240 スレッド実行。_3(4), _3(3), …とある( )内は FMM の階層番号。0 が最下層。

6. 今年度の進捗状況と今後の展望

本年度は, 基礎評価とメニーコア対応コード作成を目的として 4 期に分け研究を進める計画であった。FX100 が 4 月 1 日より使用できないとのアクシデントはあったが, 全般に順調に当初計画を進めることができた。今年度与えられた計算機資源については, FX10 を追加分を含め 100%, FX100 についてもほぼ 100%消費した。

第一期 (4-6 月) および第二期 (7-9 月) において, スレッド並列の事前評価を FX10 および Xeon Phi 上で行い, 図 3, 10 および 11 にあるようなプログラムの傾向を把握した。そののち, 第二期において, ホットスポット二箇所 (energy_direct, energy_fmm_2 および_3) についてのコード改良およびコードの高度化を行った。energy_direct については, まず図 5 に示した方法でのスレッド並列対象

iatom ループ長の伸張により 16 スレッド実行時のインバランス解消および演算性能の向上を得た。energy_fmm_2 および_3 についてはチャンクサイズを階層ごと最適化することでインバランスを低減させた。

後半の第三期および第四期では, FX100 および Xeon Phi 上においてさらなるスレッド並列性能向上に取り組んだ. Energy_direct について図 7 にあるセル単位スレッド並列化を実装することで, 32 スレッド実行時のインバランスを 1% 以下に抑えることに成功した. 結果として energy_direct オリジナルコードにくらべ 1.5 倍 (16core) および 1.7 倍 (32core) の高速化が実現された。さらに Xeon Phi において 120 スレッドまでの高速化が実現された（図 10)。本来優位のはずの改良その２の性能が芳しくない, チャンクサイズの計算式に改善の余地がある, などの問題点は残るものの 100 スレッド以上での並列性を確保したことは本研究の大きな成果である。なお energy_direct については, 改良その 1 および 2 の他にも並列化方法を複数提案し最適解を探っている最中である。

採択済みの平成 28 年度 JHPCN 課題では、引き続き Xeon Phi での性能向上, および 5 章に挙げた問題点 (ホットスポット以外でのスレッド間のバリア同期, キャッシュミス率, 整数演算の最適化など) の解決に取り組む。バリア同期削減については今回行ったと同じ改良をプログラムの該当箇所に広く適用することでプログラム全体のさらなる性能向上が期待できる。さらに開発した複数のコードについて, 自動性能チューニング技術を用い, 問題規模(原子数), プロセス数, コア数といった実行条件ごと最適なコードを探索する。また FX100 で得られた性能測定データをもとにエクサスケールマシンでの性能予測をするための数値モデルの構築も計画している。

さらに次年度 JHPCN 課題では FX100 および Xeon Phi のワイド SIMD 幅を活用できるような性能チューニングを開始する。ワイド SIMD 幅への対応は, メニーコアへの対応とともに次世代スーパーコンピュータの性能を引き出すために不可欠である。それとともに, 改良されたコードを用いた FX100 上での大規模性能テスト, および実際のサイエンス研究への応用を行う。ウィルスや高分子などを題材に, 新規サイエンス開拓を目的とした FX100 を全ノード規模利用した原子数 1000 万 ~ 1 億オーダーでの大規模 MD 計算を計画している。

7. 研究成果リスト

(1) 学術論文

該当無し

(2) 国際会議プロシーディングス

該当無し

(3) 国際会議発表

該当無し

(4) 国内会議発表

・安藤嘉倫, 吉井範行, 藤本和士, 小嶋秀和, 山田篤志, 岩橋建輔, 水谷文保, 岡崎進, 高並列対応汎用分子動力学シミュレーションソフト MODYLAS による大規模分子動力学計算, HPCS2015, 東京 (2015).

・安藤嘉倫, 吉井範行, 岡崎進, 汎用 MD ソフトウェア MODYLAS の異方的な MPI プロセス分割および基本セル分割への拡張, 第 29 回分子シミュレーション討論会, 新潟 (2015).

・安藤嘉倫, 鈴木惣一朗, 大島聡史, オーガナイズドセッション「分子動力学計算ソフトウェア MODYLAS のメニーコアアーキテクチャ対応並列化に関する研究」, HPCS2016, 仙台 (2016).

(5) その他（特許, プレス発表, 著書等）

安藤嘉倫, “汎用分子動力学計算ソフトウェア MODYLAS”, 日本機械学会 計算力学部門 CMD Newsletter, No.54, 28-30 (2015).

参考文献

[1] Y.Andoh, N.Yoshii, et al., “All-atom molecular dynamics calculation study of entire poliovirus empty capsides in solution”, *J. Chem. Phys.*, **141**, 165101 (2014).

[2] Y.Andoh, N.Yoshii, et al., “MODYLAS: A highly parallelized general-purpose molecular dynamics simulation program for large-scale systems with long-range forces calculated by fast multipole method (FMM) and highly scalable fine-grained new parallel processing algorithms”, *J. Chem. Theory Compt.*, **9**, 3201-3209 (2013).

[3] N. Yoshii, Y.Andoh, et al., “MODYLAS: A highly parallelized general-purpose molecular dynamics simulation program”, *Int. J. Quantum Chem.*, **115**, 342-348 (2015).